Electronic Devices and Circuits

IN THREE VOLUMES

VOLUME 3

by

G. J. PRIDHAM, B.SC. (ENG.), C.ENG., M.I.E.E.,
M.I.E.R.E.

Principal Lecturer, Enfield College of Technology

PERGAMON PRESS

OXFORD · NEW YORK · TORONTO
SYDNEY · BRAUNSCHWEIG

Pergamon Press Ltd., Headington Hill Hall, Oxford
Pergamon Press Inc., Maxwell House, Fairview Park, Elmsford, New York 10523
Pergamon of Canada Ltd., 207 Queen's Quay West, Toronto 1
Pergamon Press (Aust.) Pty. Ltd., 19a Boundary Street, Rushcutters Bay, N.S.W. 2011, Australia
Vieweg & Sohn GmbH, Burgplatz 1, Braunschweig

Copyright © 1972 Pergamon Press Ltd.

All Rights Reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of Pergamon Press Ltd.

First edition 1972
Library of Congress Catalog Card No. 67-26692

Printed in Hungary

This book is sold subject to the condition that it shall not, by way of trade, be lent, resold, hired out, or otherwise disposed of without the publisher's consent, in any form of binding or cover other than that in which it is published.

08 016755 1 (flexicover)
08 016626 1 (hard cover)

Introduction

THIS is the third of three volumes written for aspiring professional and technician engineers in the electronics industry. The three volumes provide a comprehensive series on electronics suitable for students on University Engineering Degree Courses and on Higher National Diploma and Certificate Courses in Electrical and Electronic Engineering.

The traditional thermionic-valve approach to electronics is no longer appropriate, and the extensive development of semiconductor devices has prompted the writing of these volumes and the use of an integrated treatment for the two devices. Introductory network theory and physics have been included to support these topics.

N. HILLER, *Editor*

List of Symbols

α	Attenuation coefficient (nepers)
A	Area (metre^2)
A	Gain of an amplifier
β	Feedback ratio
β	Phase change coefficient (radians)
B	Flux density (tesla)
B	Susceptance (siemens)
C	Capacitance (farads)
c	Velocity of light *in vacuo* (3×10^8 m/s)
δ	Width of depletion layer
D	Diffusion coefficient
D	Electric flux density (coulombs/metre^2)
ε	Base of natural logarithms
ε_0	Permittivity of free space ($1/36 \times 10^{-9}$ farads/metre)
ε_r	Relative permittivity
E	Electric field strength (volts/metre)
E	d.c. or r.m.s. value of e.m.f. (volts)
E	Energy (joules or electron volts)
E_A	Acceptor energy level (eV)
E_C	Energy level of conduction band (eV)
E_D	Donor energy level (eV)
E_F	Fermi level (eV)
E_V	Energy level of valence band (eV)
e	Charge on an electron ($1{\cdot}6 \times 10^{-19}$ C)
e	Instantaneous value of an e.m.f. (volts)
F	Force (newtons)

f	Frequency (Hz)
γ	Propagation coefficient
γ	Correlation coefficient
G	Conductance (siemens)
g_m	Mutual conductance (mA/V)
ζ	Efficiency (usually a percentage)
H	Magnetic field strength
h	Planck's constant ($6{\cdot}624 \times 10^{-34}$ joule-second)
h_f	Short-circuit current gain of a transistor
h_i	Input resistance of a transistor with output short circuited
h_o	Output conductance of a transistor with input open circuited
h_r	Voltage feedback ratio of a transistor with input open circuited
h_{FB} (or $\bar{\alpha}$)	Large signal common-base current gain
h_{FBI} (or $\bar{\alpha}_I$)	Large signal inverse common-base gain
I	d.c. or r.m.s. value of current (amp)
I	Moment of inertia
I_M	Maximum value of current (amp)
I_{C0}	Leakage current in common base (amp)
I'_{C0}	Leakage current in common emitter (amp)
i	Instantaneous value of current (amp)
$\bar{i}$	Laplace transform of current
J	Current density (amp/metre^2)
j	90° operator
K	Coefficient of coupling
k	Boltzmann's constant ($1{\cdot}38 \times 10^{-23}$ J/K)
L	Inductance (henrys)
L	Diffusion length
l	Subsidiary quantum number
μ	Amplification factor of a valve or FET
μ_0	Permeability of free space (H/m)

μ_r	Relative permeability
μ_E	Mobility of an electron (metre^2/volt-second)
μ_H	Mobility of a hole (metre^2/volt-second)
M	Momentum
M	Mutual inductance (henrys)
m	Magnetic quantum number
m_e	Effective mass of an electron ($9{\cdot}1 \times 10^{-31}$ kg)
m_h	Effective mass of a hole
m_s	Spin quantum number
n	Principal quantum number
n	Number of turns
N_A	Number of acceptor impurity atoms per cubic metre
N_D	Number of donor impurity atoms per cubic metre
N_E	Electron density
N_H	Hole density
N_I	Number of charge carriers per cubic metre in an intrinsic semiconductor
P	Power (watts)
Q	Charge (coulombs)
q	Instantaneous value of charge (coulombs)
ϱ	Resistivity (Ω-m)
ϱ	Charge density (C/m^3)
R	Resistance (Ω)
R_D	Dynamic resistance (Ω)
r	Radius (m)
r_e, r_b, r_c	T parameters
$r_{b'e}$, $r_{b'b}$, r_{ce}, $r_{cb'}$	Hybrid-π parameters
r_a	Anode slope resistance
r_{ds}	Drain resistance
θ	Hall angle
s	Laplace operator

T	Absolute temperature (K)
t	Time (seconds)
Φ	Magnetic flux (webers)
ϕ	Instantaneous value of magnetic flux (webers)
ϕ	Potential barrier
V	d.c. or r.m.s. voltage (volts)
v	Maximum value of voltage (volts)
v	Instantaneous value of voltage (volts)
$\boldsymbol{v}$	Velocity (m/s)
X	Reactance (Ω)
Y	Admittance ($= 1/Z$) (siemens)
y_i	Input admittance with output short circuited
y_o	Output admittance with input short circuited
y_f	Forward transfer admittance with output short circuited
y_r	Reverse transfer admittance with input short circuited
ψ	Electric flux (coulombs)
Z	Impedance (Ω)
Z_0	Characteristic impedance (Ω)
ω	Angular frequency (rads/s)

Where subscripts are used, upper case denotes d.c. values and lower case r.m.s. values.

CHAPTER 1

Network Theory

IN THIS chapter the use of Laplace transforms for network analysis will be discussed and filter networks analysed in more detail than in the previous volumes.

1.1. Laplace Transforms and Their Applications

The Laplace transform of any function is defined by the relation

$$\mathcal{L}[F(t)] = \int_0^\infty \varepsilon^{-st} F(t)\, dt,$$

and once the differential equation(s) for any circuit has been set up, a method of solution is to take the transforms of each side, separate into partial fractions, and then evaluate the inverse of the transforms.

It is not the purpose of this book to cover material that rightly belongs to a mathematical treatise, but in the following examples an indication of the mathematical background will be given.

1.1.1. Simple RL circuit with suddenly applied d.c.

For the circuit shown in Fig. 1.1, at any time t after the switch is closed,

$$L\frac{di}{dt}+Ri = V.$$

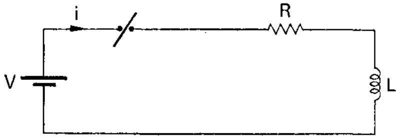


FIG. 1.1. Simple *RL* circuit.

The transform of $\dfrac{di}{dt}$ is $-i_0+s\bar{i}$.

The transform of i is $\bar{i}$.

The transform of V is $\dfrac{V}{s}$.

Hence, taking transforms of both sides,

$$L(-i_0+s\bar{i})+R\bar{i} = \frac{V}{s}.$$

If, at time $t = 0$, $i = 0$, then $i_0 = 0$.

Therefore $$(Ls+R)\bar{i} = \frac{V}{s},$$

i.e. $$\bar{i} = \frac{V}{Ls[s+(R/L)]}$$
$$= \frac{V}{L}\left(\frac{L/R}{s}-\frac{L/R}{s+(R/L)}\right)$$
$$= \frac{V}{R}\left(\frac{1}{s}-\frac{1}{s+(R/L)}\right).$$

The inverse of $\dfrac{1}{s}$ is 1.

The inverse of $\dfrac{1}{[s+(R/L)]}$ is $\varepsilon^{-(Rt/L)}$.

Hence $$i = \frac{V}{R}(1-\varepsilon^{(Rt/L)}).$$

This equation gives the current at any time t after the switch is closed.

1.1.2. Series LCR circuit with suddenly applied d.c.

At any time t after the direct voltage V is applied, the equations for the series LCR circuit shown in Fig. 1.2 are

$$L\frac{di}{dt}+Ri+\frac{q}{C} = V \quad \text{and} \quad i = \frac{dq}{dt}.$$

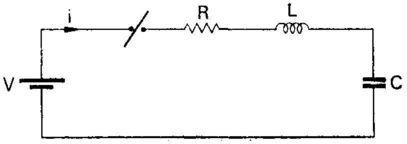


FIG. 1.2. *RLC* circuit.

The transforms of these equations are

$$L(-i_0+s\bar{i})+R\bar{i}+\frac{\bar{q}}{C} = \frac{V}{s}$$

and $$\bar{i} = -q_0+s\bar{q}.$$

If at time $t = 0$, $q = i = 0$, then $i_0 = q_0 = 0$.

Hence $$(Ls+R)\bar{i}+\frac{\bar{q}}{C} = \frac{V}{s}$$

and $$\bar{i} = s\bar{q}.$$

Eliminating $\bar{q}$ between these expressions gives

$$\left(Ls+R+\frac{1}{Cs}\right)\bar{i}=\frac{V}{s}$$

or

$$\bar{i}=\frac{V}{L[s^2+(R/L)s+(1/LC)]}=\frac{V}{L\{[s+(R/2L)]^2+[(1/LC)-(R^2/4L^2)]\}}\,.$$

The form of the inverse of this transform depends on the relative values of L, C, and R.

(i) If $1/LC > R^2/4L^2$, the transform is of the form $k/[(s+\alpha)^2+\beta^2]$ and has an inverse given by $(k/\beta)\,\varepsilon^{-\alpha t}\sin\beta t$.

The actual value of the inverse is

$$i=\frac{V}{L\sqrt{[(1/LC)-(R^2/4L^2)]}}\,\varepsilon^{-(Rt/2L)}\sin\left[\sqrt{\left(\frac{1}{LC}-\frac{R^2}{4L^2}\right)}t\right]$$

and indicates that the response is a sinusoidal oscillation of gradually decreasing amplitude.

(ii) If $1/LC = R^2/4L^2$, the transform may be written

$$\bar{i}=\frac{V}{L[s+(R/2L)]^2}$$

and the inverse is given by

$$i=\frac{V}{L}\,t\varepsilon^{-(Rt/2L)}$$

and indicates a critically damped response.

(iii) When $1/LC < R^2/4L^2$, $\bar{i}$ is of the form $k/[(s+\alpha)^2-\beta^2]$, and the inverse of this is $(k/\beta)\varepsilon^{-\alpha t}\sinh\beta t$, i.e. an overdamped response.

Under this condition the current is given by

$$i=\frac{V}{L\sqrt{[(R^2/4L^2)-(1/LC)]}}\varepsilon^{-(Rt/2L)}\sinh\left[\sqrt{\left(\frac{R^2}{4L^2}-\frac{1}{LC}\right)}t\right].$$

1.1.3. Simple RC circuit with suddenly applied a.c.

The equation for the circuit shown in Fig. 1.3, in which C is initially uncharged, is

$$iR+\frac{1}{C}\int_0^t i\,dt=V_m\sin\omega t.$$

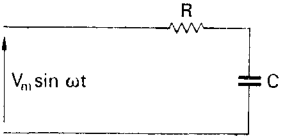


FIG. 1.3. Simple *RC* circuit.

The transform of $\int_0^t i\,dt$ is $\frac{\bar{i}}{s}$.

The transform of $\sin\omega t$ is $\frac{\omega}{s^2+\omega^2}$.

Hence, taking transforms of the above equation,

$$R\bar{i}+\frac{\bar{i}}{Cs}=\frac{\omega V_m}{s^2+\omega^2}$$

or

$$\bar{i}=\frac{\omega V_m s}{R[s+(1/RC)](s^2+\omega^2)}$$

$$=\frac{\omega V_m}{R}\left[\frac{A}{[s+(1/RC)]}+\frac{Bs+D}{(s^2+\omega^2)}\right],$$

where A, B, and D are constants.

But

$$s \equiv A(s^2+\omega^2)+(Bs+D)\left(s+\frac{1}{RC}\right).$$

Equating coefficient of s^2;

$$0 = A+B.$$

Equating coefficients of s;

$$1 = D+\frac{B}{RC}.$$

Equating constants;

$$0 = A\omega^2+\frac{D}{RC}.$$

From these relations,

$$1 = -A\omega^2RC-\frac{A}{RC}.$$

Therefore

$$A = -\frac{RC}{1+\omega^2R^2C^2},$$

$$B = \frac{RC}{1+\omega^2R^2C^2},$$

and

$$D = \frac{\omega^2R^2C^2}{1+\omega^2R^2C^2};$$

i.e.

$$\bar{i} = \frac{\omega V_m}{R}\left[-\frac{RC}{(1+\omega^2R^2C^2)\,[s+(1/RC)]}\right.$$
$$\left.+\left(\frac{RCs}{1+\omega^2R^2C^2}+\frac{\omega^2R^2C^2}{1+\omega^2R^2C^2}\right)\left(\frac{1}{s^2+\omega^2}\right)\right]$$
$$= \frac{\omega V_mC}{(1+\omega^2R^2C^2)}\left[-\frac{1}{[s+(1/RC)]}+\frac{s+\omega^2RC}{(s^2+\omega^2)}\right].$$

The inverse of $\dfrac{1}{s+(1/RC)}$ is $\varepsilon^{-(t/RC)}$.

The inverse of $\dfrac{s}{s^2+\omega^2}$ is $\cos \omega t$.

The inverse of $\dfrac{\omega}{s^2+\omega^2}$ is $\sin \omega t$.

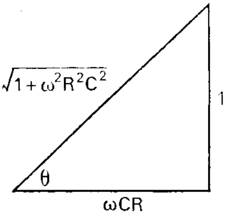


FIG. 1.4. Phase angle for *RC* circuit.

Hence, taking the inverse of the above expression,

$$i = \frac{\omega V_m C}{1+\omega^2R^2C^2}\,[\varepsilon^{-(t/RC)} + \cos \omega t + \omega RC \sin \omega t].$$

If $\dfrac{1}{\omega RC} = \tan \theta$ (see Fig. 1.4),

$$i = \frac{\omega V_m C}{\sqrt{(1+\omega^2R^2C^2)}}\,[-\varepsilon^{-(t/RC)} \sin \theta + \cos \omega t \sin \theta + \sin \omega t + \cos \theta]$$

$$= \frac{\omega V_m C}{\sqrt{(1+\omega^2R^2C^2)}}\,[-\varepsilon^{-(t/RC)} \sin \theta + \sin (\omega t + \theta)]$$

$$= -\frac{V_m\,\varepsilon^{-(t/RC)} \sin \theta}{\sqrt{[R^2+(1/\omega^2C^2)]}} + \frac{V_m \sin (\omega t + \theta)}{\sqrt{[R^2+(1/\omega^2C^2)]}}.$$

The first term in this expression represents the transient term which disappears as *t* increases, while the second represents the steady-state condition.

1.2. Filter Networks

Two port networks were introduced in Chapter 1 of Volume 2, and the main parameters evaluated. For the symmetrical *T* section shown in Fig. 1.5, the characteristic impedance Z_{0T} defined as the input impedance of an infinite number of cascaded sections, is given by

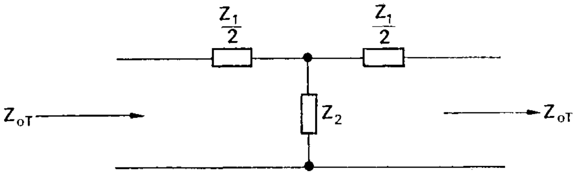


FIG. 1.5. Symmetrical *T* network.

$$Z_{0T} = \frac{Z_1}{2} + \frac{Z_2[(Z_1/2)+Z_{0T}]}{(Z_1/2)+Z_2+Z_{0T}},$$

which leads to

$$Z_{0T} = \sqrt{\left[\left(\frac{Z_1}{2}\right)^2 + Z_1Z_2\right]}.$$

Also the propagation coefficient γ defined as the natural logarithm of the complex ratio of the steady-state input and output currents of a symmetrical section in an infinite number of such sections, is given by reference to Fig. 1.6,

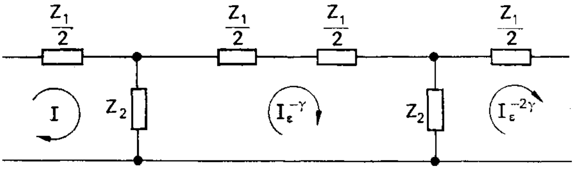


FIG. 1.6. Propagation coefficient of *T* network.

i.e. $$I\varepsilon^{-\gamma}(Z_1+2Z_2) - IZ_2 - I\varepsilon^{-2\gamma}Z_2 = 0,$$

which leads to

$$\cosh \gamma = 1+\frac{Z_1}{2Z_2}.$$

If the symmetrical π section shown in Fig. 1.7 is considered the characteristic admittance is given by

$$Y_{0\pi} = \frac{Y_2}{2}+\frac{Y_1[Y_{0\pi}+(Y_2/2)]}{Y_1+(Y_2/2)+Y_{0\pi}},$$

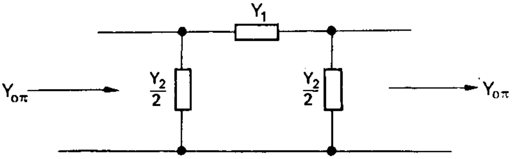


FIG. 1.7. Symmetrical π network.

which leads to

$$Y_{0\pi} = \sqrt{\left[\left(\frac{Y_2}{2}\right)^2+Y_1Y_2\right]}$$

and
$$Z_{0\pi} = \frac{1}{Y_{0\pi}} = \frac{Z_1Z_2}{Z_{0\pi}}.$$

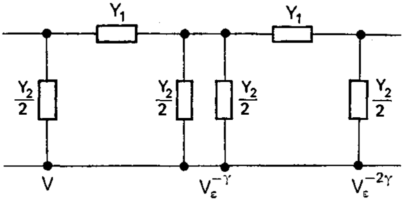


FIG. 1.8. Propagation coefficient of π network.

Also, by reference to Fig. 1.8,

$$(V-V\varepsilon^{-\gamma})Y_1 = V\varepsilon^{-\gamma}Y_2+(V\varepsilon^{-\gamma}-V\varepsilon^{-2\gamma})Y_1$$

or
$$\cosh\gamma = 1+\frac{Y_2}{2Y_1}$$
$$= 1+\frac{Z_1}{2Z_2}.$$

Hence for both sections the propagation coefficient γ is identical and if γ is written in the form $\alpha+j\beta$, where α is the attenuation coefficient and β the phase change coefficient.

$$\cosh(\alpha+j\beta) = 1+\frac{Z_1}{2Z_2}.$$

Expanding and replacing $\cosh j\beta$ by $\cos\beta$ and $\sinh j\beta$ by $j\sin\beta$,

$$\cosh\alpha\cos\beta+j\sinh\alpha\sin\beta = 1+\frac{Z_1}{2Z_2}.$$

In ideal filter networks both Z_1 and Z_2 are pure reactances, and if real terms are equated,

$$\cosh\alpha\cos\beta = 1+\frac{Z_1}{2Z_2}. \qquad (1.2.1)$$

The j terms give

$$\sinh\alpha\sin\beta = 0. \qquad (1.2.2)$$

Equation (1.2.2) leads to two possible solutions:

β finite $\quad \sinh\alpha = 0,\quad$ i.e. $\quad \alpha = 0,$

α finite $\quad \sin\beta = 0,\quad$ i.e. $\quad \beta = n\pi,$

where n is an integer.

The condition $\sinh\alpha = 0$ and β finite gives the condition for zero attenuation, i.e. the pass band of a filter; while $\sin\beta = 0$ and α finite gives the condition for the attenuation band of a filter.

The pass-band condition, i.e. $\sinh\alpha = 0$, means that $\cosh\alpha = 1$, and in the pass band

$$\cos\beta = 1+\frac{Z_1}{2Z_2}\,.$$

1.3. Prototype (Constant-*k*) Filters

These filters may be high-pass, low-pass, band-pass, or band-stop, and the term constant-k simply means that the product of the series and shunt arms impedances is a constant, i.e. $Z_1Z_2 = k^2$, where k is a real constant.

1.3.1. Prototype low-pass filters

This type of filter is shown in Fig. 1.9 (a) and (b) with

$$Z_1 = j\omega L \quad \text{and} \quad Z_2 = \frac{1}{j\omega C}\,.$$

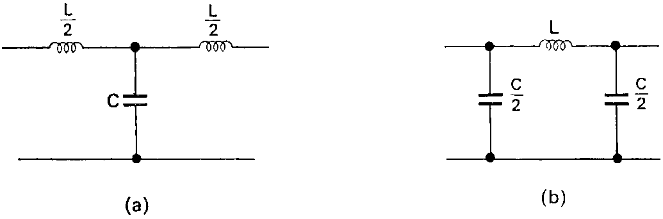


FIG. 1.9. Constant-*k* low-pass filters.

Then, in the pass band,

$$\cos\beta = 1-\frac{1}{2}\omega^2 LC \quad \text{or} \quad \omega^2 = \frac{2}{LC}(1-\cos\beta).$$

The limits of $\cos\beta$ are ±1, and hence the limits of ω are 0 and $2/\sqrt{(LC)}$,

i.e. $$f = 0 \quad \text{and} \quad \frac{1}{\pi\sqrt{(LC)}}\,.$$

This shows that such a filter will pass frequencies from zero to a cut-off value of

$$f_c = \frac{1}{\pi\sqrt{(LC)}}.$$

Within this pass band,

$$\begin{aligned}\cos\beta &= 1+\frac{Z_1}{2Z_2}\\ &= 1-\frac{\omega^2 LC}{2}\\ &= 1-2\left(\frac{\omega}{\omega_c}\right)^2,\end{aligned}$$

i.e.

$$1-2\sin^2\frac{\beta}{2} = 1-2\left(\frac{\omega}{\omega_c}\right)^2,$$

i.e.

$$\beta = 2\sin^{-1}\pm\left(\frac{\omega}{\omega_c}\right).$$

This shows that within the pass band the angle by which the output current lags or leads the input current follows an inverse sine relationship if the section is one of an infinite number of sections.

In the attenuation band when $\beta = n\pi, \cos\beta = \pm 1$,

i.e.

$$\pm\cosh\alpha = 1-\tfrac{1}{2}\omega^2 LC$$

or

$$\cosh\alpha = \pm\left[1-2\left(\frac{\omega}{\omega_c}\right)^2\right].$$

Since ω is greater than ω_c, the negative sign must be taken and, putting $\cosh\alpha = 2\cosh^2\frac{\alpha}{2}-1$,

$$2\cosh^2\frac{\alpha}{2}-1 = 2\left(\frac{\omega}{\omega_c}\right)^2-1,$$

i.e.

$$\alpha = 2\cosh^{-1}\left(\frac{\omega}{\omega_c}\right) \quad \text{nepers.}$$

This shows that the attenuation follows an inverse cosh relationship as indicated by Fig. 1.10.

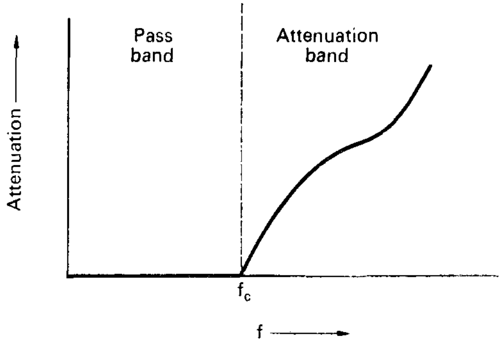


FIG. 1.10. Attenuation coefficient of constant-k low-pass filter.

Also the fact that the negative sign has been chosen for cosh α means that $\beta = \pm\pi$. The positive sign must be chosen for β as indicated by Fig. 1.11. At high frequencies, above cut-off, the circuit is inductive and the voltage across the capacitor V_c must lag the input current by 90°. This is the drive for the output current which lags V_c by a further 90°. Hence the output current lags the input current by π or $\beta = \pi$.

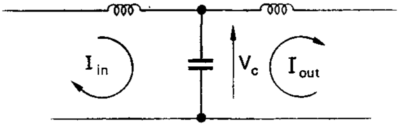


FIG. 1.11. Currents in a constant-k low-pass filter.

The complete variation of the phase angle of output current with respect to the input current is shown in Fig. 1.12.

The characteristic impedance of the T-type, low-pass section is given by

$$Z_{0T} = \sqrt{\left[\left(\frac{Z_1}{2}\right)^2 + Z_1Z_2\right]},$$

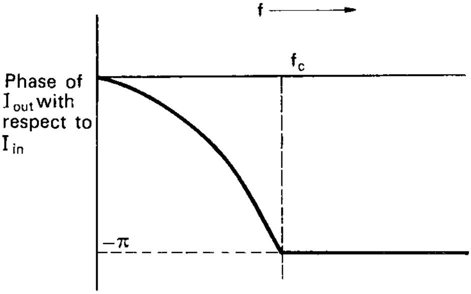


FIG. 1.12. Phase relation between input and output current of a low-pass filter.

where $$Z_1 = j\omega L, \quad Z_2 = \frac{1}{j\omega C},$$

i.e. $$Z_{0T} = \sqrt{\left[-\frac{\omega^2 L^2}{4} + \frac{L}{C}\right]}$$

$$= \sqrt{\frac{L}{C}}\sqrt{\left[1 - \frac{\omega^2 LC}{4}\right]}$$

$$= \sqrt{\frac{L}{C}}\sqrt{\left[1 - \left(\frac{\omega}{\omega_c}\right)^2\right]}.$$

This obviously varies with frequency. At very low frequencies $Z_{0T} = \sqrt{\dfrac{L}{C}}$ falling to zero at $\omega = \omega_c$ and becoming inductive above cut-off.

For the π section,

$$Z_{0\pi} = \frac{Z_1 Z_2}{Z_{0T}} = \frac{L/C}{\sqrt{\frac{L}{C}}\sqrt{\left[1 - \left(\frac{\omega}{\omega_c}\right)^2\right]}}$$

$$= \frac{\sqrt{(L/C)}}{\sqrt{\left[1 - \left(\frac{\omega}{\omega_c}\right)^2\right]}}.$$

The characteristic impedance is also $\sqrt{(L/C)}$ at very low frequencies rising to infinity at the cut-off frequency and becoming capacitive above the cut-off frequency. The complete variations in characteristic impedance of prototype low-pass filters is shown in Fig. 1.13. In practice some blurring of these curves occurs near the cut-off frequency due to the inherent resistance of the inductor.

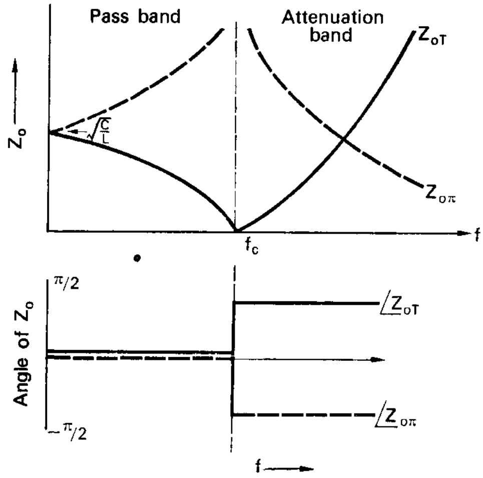


FIG. 1.13. Characteristic impedance of constant-*k* low-pass filters.

When the filter is incorporated in a system, correct matching can only be obtained at one particular frequency. This is chosen at the geometric mean frequency of the pass band, i.e. at a frequency, $f = \sqrt{(0 \times f_c)} = 0$.

Then the design impedance R_0, i.e. the characteristic impedance at zero frequency is given by $R = \sqrt{(L/C)}$. Hence since $f_c = \dfrac{1}{\pi\sqrt{(L/C)}}$, the values of series inductance and shunt capacitance are given by

$$L = \frac{R_0}{\pi f_c} \quad \text{and} \quad C = \frac{1}{\pi f_c R_0}.$$

If a particular case is chosen, e.g. a low-pass filter, having a cut-off frequency of 10 kHz in a 75 Ω system,

$$L = \frac{75}{\pi \times 10^4}\,\text{H} = 2{\cdot}39\ \text{mH},$$

$$C = \frac{1}{\pi \times 10^4 \times 75}\,\text{F} = 0{\cdot}424\ \mu\text{F}.$$

The prototype T and π low-pass filters are then as shown in Fig. 1.14 (a) and (b).

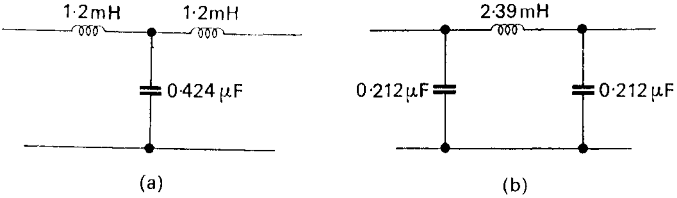


FIG. 1.14. 10 kHz low-pass filters.

1.3.2. *Prototype high-pass filters*

This type of filter is shown in Fig. 1.15 (a) and (b) with $Z_1 = \frac{1}{j\omega C}$ and $Z_2 = j\omega L$.

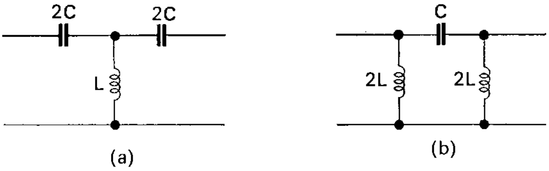


FIG. 1.15. Constant-k high-pass filters.

Then, in the pass band,

$$\cos\beta = 1-\frac{1}{2\omega^2 LC}$$

or

$$\omega^2 = \frac{1}{2LC(1-\cos\beta)}\,.$$

The limits of cos β are ± 1, and hence the limits of ω are $\omega_0 = \dfrac{1}{2\sqrt{(LC)}}$ and ∞,

i.e.

$$f_c = \frac{1}{4\pi\sqrt{(LC)}} \quad \text{and} \quad \infty.$$

This shows that such a filter will pass frequencies between a cut-off frequency $f_c = \dfrac{1}{4\pi\sqrt{(LC)}}$ and ∞.

Within this pass band,

$$\begin{aligned}\cos\beta &= 1+\frac{Z_1}{2Z_2}\\ &= 1-\frac{1}{2\omega^2 LC}\\ &= 1-2\left(\frac{\omega_c}{\omega}\right)^2,\end{aligned}$$

i.e.

$$1-2\sin^2\frac{\beta}{2} = 1-2\left(\frac{\omega_c}{\omega}\right)^2$$

or

$$\beta = 2\sin^{-1}\pm\left(\frac{\omega_c}{\omega}\right).$$

This shows that within the pass band the angle by which the output current lags or leads the input current, follows an inverse sine relationship if the section is one of an infinite number of sections.

In the attenuation band when $\beta = n\pi$, $\cos\beta = \pm 1$,

i.e.
$$\pm\cosh\alpha = 1-\frac{1}{2\omega^2 LC}$$

or
$$\cosh\alpha = \pm\left[1-2\left(\frac{\omega_c}{\omega}\right)^2\right].$$

Since ω_c is greater than ω, the negative sign must be taken, and putting

$$\cosh\alpha = 2\cosh^2\frac{\alpha}{2}-1,$$

$$2\cosh^2\frac{\alpha}{2}-1 = 2\left(\frac{\omega_c}{\omega}\right)^2-1,$$

i.e.
$$\alpha = 2\cosh^{-1}\left(\frac{\omega_c}{\omega}\right).$$

This shows that the attenuation follows an inverse cosh relationship as indicated by Fig. 1.16.

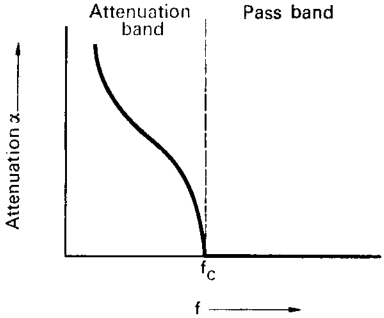


FIG. 1.16. Attenuation coefficient of constant-*k* high-pass filters.

Also, since the negative sign for cosh α has been chosen, $\beta = \pm\pi$. In fact the negative sign for β must be chosen as indicated by Fig. 1.17. At low frequencies, below cut-off, the circuit is capacitive and the voltage across the inductor V_l must lead the input current by 90°.

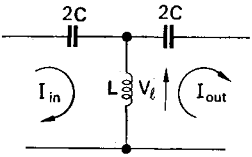


FIG. 1.17. Currents in a constant-k high-pass filter.

This is the drive for the output current which leads V_l by a further 90°. Hence $\beta = -\pi$. The complete variation of the phase of the output current with respect to the input current is shown in Fig. 1.18.

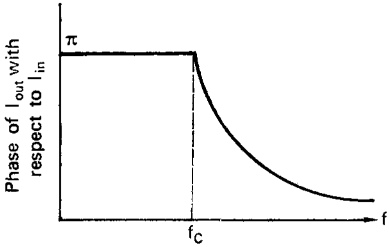


FIG. 1.18. Phase relation between input and output currents of a high-pass filter.

The characteristic impedance of the T-type, high-pass section is given by

$$Z_{0T} = \sqrt{\left[\left(\frac{Z_1}{2}\right)^2 + Z_1 Z_2\right]},$$

where $$Z_1 = \frac{1}{j\omega C}, \quad Z_2 = j\omega L,$$

i.e. $$Z_{0T} = \sqrt{\left[-\frac{1}{4\omega^2 C} + \frac{L}{C}\right]}$$

$$= \sqrt{\frac{L}{C}}\sqrt{\left[1 - \frac{1}{4\omega^2 LC}\right]}$$

$$= \sqrt{\frac{L}{C}}\sqrt{\left[1 - \left(\frac{\omega_c}{\omega}\right)^2\right]}.$$

This obviously varies with frequency. At very high frequencies $Z_{0T} = \sqrt{(L/C)}$, falling to zero at $\omega = \omega_c$ and becoming inductive below cut-off.

For the π section,

$$Z_{0\pi} = \frac{Z_1 Z_2}{Z_{0T}} = \frac{L/C}{\sqrt{\frac{L}{C}}\sqrt{\left[1-\left(\frac{\omega_c}{\omega}\right)^2\right]}}$$

$$= \frac{\sqrt{(L/C)}}{\sqrt{\left[1-\left(\frac{\omega_c}{\omega}\right)^2\right]}}.$$

The characteristic impedance is also $\sqrt{(L/C)}$ at very high frequencies, rising to infinity at the cut-off frequency and becoming capacitive below cut-off. The complete variation in characteristic impedance of prototype high-pass filters is shown in Fig. 1.19. As in the case of low-pass filters, some blurring of these curves occurs near the cut-off frequency due to the inherent resistance of the inductor.

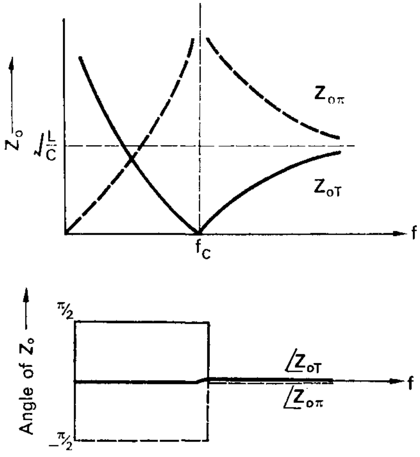


FIG. 1.19. Characteristic impedance of a high-pass filter.

When the filter is incorporated in a system, correct matching can only be obtained at one particular frequency. As before, this is chosen at the geometric mean frequency of the pass band; i.e. at a frequency,

$$f = \sqrt{f_c \infty} = \infty .$$

Then the design impedance R_0, i.e. the characteristic impedance at very high frequencies, is given by $R_0 = \sqrt{(L/C)}$. Hence, since $f_c = \dfrac{1}{4\pi\sqrt{(LC)}}$, the values of series capacitance and shunt inductance are given by

$$C = \frac{1}{4\pi R_0 f_c} \quad \text{and} \quad L = \frac{R_0}{4\pi f_c} .$$

Again, choosing a particular case, i.e. a high-pass prototype filter, having a 10 kHz cut-off frequency in a 75 Ω system.

$$C = \frac{1}{4\pi \times 75 \times 10^4}\,\text{F} = 0{\cdot}106\ \mu\text{F},$$

$$L = \frac{75}{4\pi \times 10^4}\,\text{H} = 0{\cdot}597\ \text{mH}.$$

The prototype T and π high-pass filters are then as shown in Fig. 1.20 (a) and (b).

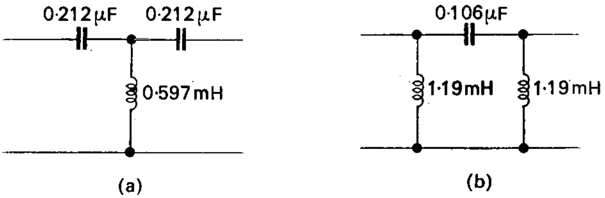


FIG. 1.20. 10 kHz high-pass filter.

1.3.3. Prototype band-pass filters

In this type of filter a low impedance series path and high impedance shunt path is provided for the required band and a low impedance shunt path and high impedance series path for unwanted frequencies. Series- and parallel-tuned circuits are used as indicated in Fig. 1.21.

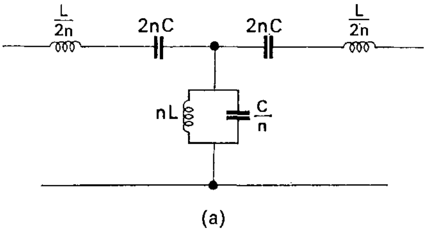
In these circuits, different values of series and shunt, inductance and capacitance, are used to control the limits of the pass band, but they must all be resonant at the same frequency. Hence the series inductance is decreased by a factor n and the series capacitance is increased by a factor n while the shunt inductance is increased and the shunt capacitance decreased by the same ratio.

Then

$$\begin{aligned} Z_1 &= \frac{j\omega L}{n} + \frac{1}{j\omega nC} \\ &= \frac{j\omega L}{n}\left(1 - \frac{1}{\omega^2 LC}\right) \\ &= \frac{j\omega L}{n}\left[1 - \left(\frac{\omega_0}{\omega}\right)^2\right], \end{aligned}$$

where $$\omega_0^2 = \frac{1}{LC}$$

and $$\begin{aligned} Z_2 &= \frac{j\omega nL \times \dfrac{n}{j\omega C}}{j\omega nL + \dfrac{n}{j\omega C}} \\ &= \frac{j\omega nL}{1 - \omega^2 LC} \\ &= \frac{j\omega nL}{\left[1 - \left(\dfrac{\omega}{\omega_0}\right)^2\right]}. \end{aligned}$$



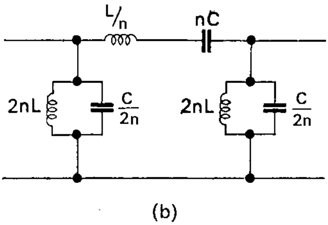


FIG. 1.21. Constant-*k* band-pass filter.

Then in the pass band,

$$\cos\beta = 1+\frac{Z_1}{2Z_2}$$

$$= 1+\frac{\left[1-\left(\frac{\omega}{\omega_0}\right)^2\right]\left[1-\left(\frac{\omega_0}{\omega}\right)^2\right]}{2n^2},$$

i.e.
$$\left[1-\left(\frac{\omega}{\omega_0}\right)^2\right]\left[1-\left(\frac{\omega_0}{\omega}\right)^2\right] = -2n^2(1-\cos\beta).$$

If $\cos\beta = 1, \omega = \omega_0$, i.e. corresponding to the resonant frequency. If

$$\cos\beta = -1, \quad \left[1-\left(\frac{\omega}{\omega_0}\right)^2\right]\left[1-\left(\frac{\omega_0}{\omega}\right)^2\right] = -4n^2,$$

i.e.
$$\left[1-\left(\frac{\omega}{\omega_0}\right)^2\right]^2 = 4n^2\left(\frac{\omega}{\omega_0}\right)^2,$$

i.e.
$$1-\left(\frac{\omega}{\omega_0}\right)^2 = \pm 2n\left(\frac{\omega}{\omega_0}\right)$$

or
$$\left(\frac{\omega}{\omega_0}\right)^2 \pm 2n\left(\frac{\omega}{\omega_0}\right) - 1 = 0,$$

i.e.
$$\left(\frac{\omega}{\omega_0}\right) = \pm n \pm \sqrt{(n^2+1)}.$$

The possible solutions are

$$\left(\frac{\omega}{\omega_0}\right) = \sqrt{(n^2+1)} \pm n.$$

Hence the limits of the pass band are

$$f_1 = f_0[\sqrt{(n^2+1)}+n] \quad \text{and} \quad f_2 = f_0[\sqrt{(n^2+1)}-n],$$

and the pass band covers a frequency range $2nf_0$. Within this pass band,

$$\cos\beta = 1+\frac{Z_1}{2Z_2}$$
$$= 1-\frac{(\omega^2-\omega_0^2)^2}{2n^2\omega^2\omega_0^2}$$

or
$$1-2\sin^2\frac{\beta}{2} = 1-2\left(\frac{\omega^2-\omega_0^2}{2n\omega\omega_0}\right),$$

i.e.
$$\beta = 2\sin^{-1}\pm\left(\frac{\omega\ -\omega_0^2}{2n\omega\omega_0}\right).$$

As in the case of low- and high-pass filters, this shows that the angle by which the output current lags or leads the input current follows an inverse sine relationship if the section is one of an infinite number of sections.

In the attenuation band, when $\beta = n\pi$, $\cos\beta = \pm 1$,

i.e.
$$\pm\cosh\alpha = 1-\frac{(\omega^2-\omega_0^2)^2}{2n^2\omega^2\omega_0^2},$$

i.e.
$$\cosh\alpha = \pm\left[1-2\left(\frac{\omega^2-\omega_0^2}{2n\omega\omega_0}\right)^2\right].$$

Cosh α must be positive, and hence the negative sign must be taken.

Then

$$2\cosh^2\frac{\alpha}{2}-1 = 2\left(\frac{\omega^2-\omega_0^2}{2n\omega\omega_0}\right)-1$$

or
$$\alpha = 2\cosh^{-1}\pm\left(\frac{\omega-\omega_0}{2n\omega\omega_0}\right).$$

This shows that outside the pass band the attenuation follows an inverse cosh relationship, the positive and negative signs being chosen for frequencies above and below the pass band respectively. The complete relationship is shown in Fig. 1.22.

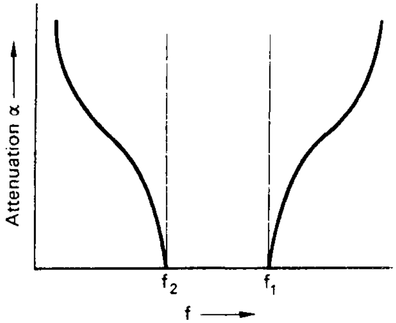


Fig. 1.22. Attenuation coefficient of a band-pass filter.

Below the pass band, when ω is less than ω_0, the positive sign must be chosen in the expression $\beta = 2\sin^{-1}\pm((\omega^2-\omega_0^2)/(2n\omega\omega_0))$, and by a similar reasoning to that employed for the high-pass filter the output current will lead the input current by π. Above the pass band (ω greater than ω_0) the output current will lag the input current by π. The complete variation in the phase of the output current with respect to the input current is shown in Fig. 1.23.

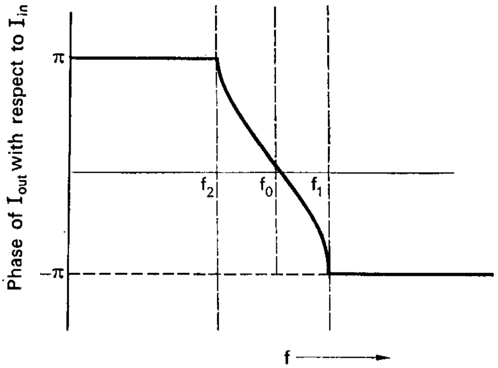


FIG. 1.23. Phase relation between input and output currents of a band-pass filter.

The characteristic impedance of the T-type band-pass section is given by

$$Z_{0T} = \sqrt{\left[\left(\frac{Z_1}{2}\right)^2 + Z_1 Z_2\right]},$$

where

$$Z_1 = \frac{j\omega L}{n}\left[1-\left(\frac{\omega_0}{\omega}\right)^2\right] \quad \text{and} \quad Z_2 = \frac{j\omega n L}{\left[1-\left(\frac{\omega}{\omega_0}\right)^2\right]}.$$

Therefore

$$Z_{0T} = \sqrt{\left\{-\frac{\omega^2 L^2}{n^2}\left[1-\left(\frac{\omega_0}{\omega}\right)^2\right]^2 + \omega_0^2 L^2\right\}}$$

$$= \sqrt{\left\{-\frac{\omega^2 L^2}{n^2}\left[1-\left(\frac{\omega_0}{\omega}\right)^2\right]^2 + \frac{L}{C}\right\}} \quad \text{since} \quad \omega_0^2 = \frac{1}{LC}.$$

At the resonant frequency ω_0, $Z_{0T} = \sqrt{(L/C)}$, becoming zero at the limits of the pass band and inductive in the attenuation band.

For the π section the characteristic impedance is given by

$$Z_{0\pi} = \frac{Z_1 Z_2}{Z_{0T}}$$

$$= \frac{\sqrt{(L/C)}}{\sqrt{\left\{1 - \frac{\omega^2}{\omega_0^2 n^2}\left[1 - \left(\frac{\omega_0}{\omega}\right)^2\right]^2\right\}}}.$$

This is also $\sqrt{(L/C)}$ at the resonant frequency rising to infinity at the limits of the pass band and becoming capacitive in the attenuation band. The complete variation in characteristic impedance of prototype band-pass filters is shown in Fig. 1.24.

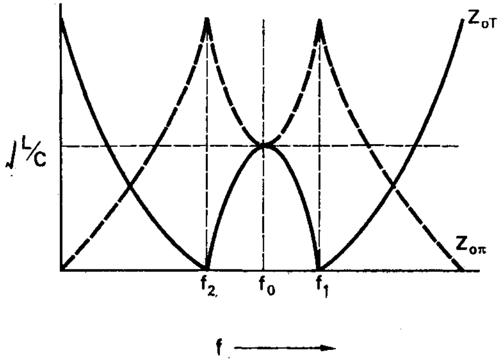


FIG. 1.24. Characteristic impedance of a constant-k band-pass filter.

As in previous cases, correct matching of the filter can only be obtained at one particular frequency chosen at the geometric mean of the pass band,

i.e. $$f = \sqrt{\{f_0[\sqrt{(n^2+1)}+n]\, f_0[\sqrt{(n^2+1)}-n]\}} = f_0.$$

Then the design impedance R_0, i.e. the characteristic impedance at f_0 is given by $R_0 = \sqrt{(L/C)}$. Hence, since $f_0 = 1/[2\pi\sqrt{(LC)}]$ the values of L and C are given by $L = R_0/2\pi f_0$ and $C = 1/2\pi f_0 R_0$, where f_0 is the geometric mean frequency of the pass band. The series and shunt elements may then be determined by putting

$$n = \frac{f_1 - f_2}{2f_0}.$$

As an example of this type of filter, consider a prototype band-pass filter for the range 10 kHz to 12·1 kHz in a 75 Ω system.

$$f_0 = \sqrt{(10 \times 12{\cdot}1)} = 11 \text{ kHz}.$$

Then

$$L = \frac{75}{2\pi \times 11 \times 10^3} \text{ H} = 1{\cdot}06 \text{ mH},$$

$$C = \frac{10^6}{2\pi \times 11 \times 10^3 \times 75} = 0{\cdot}193\ \mu\text{F},$$

$$n = \frac{2{\cdot}1 \times 10^3}{2 \times 11 \times 10^3} = 0{\cdot}095,$$

i.e. the series inductance is $L/n = 11{\cdot}2$ mH;
the series capacitance is $nC = 0{\cdot}0183\ \mu F$;
the shunt inductance is $nL = 0{\cdot}101$ mH;
the shunt capacitance is $C/n = 2{\cdot}03\ \mu$F.

Complete prototype T and π band-pass filters are shown in Fig. 1.25.

1.3.4. Prototype band-stop filters

In this type of filter a high impedance series path and low impedance shunt path is provided for the band that is to be stopped, and a low impedance series path and high impedance shunt path for wanted frequencies. Parallel- and series-tuned circuits are used as indicated in Fig. 1.26.

In these circuits, different values of series and shunt, inductance and capacitance are used to control the limits of the band that

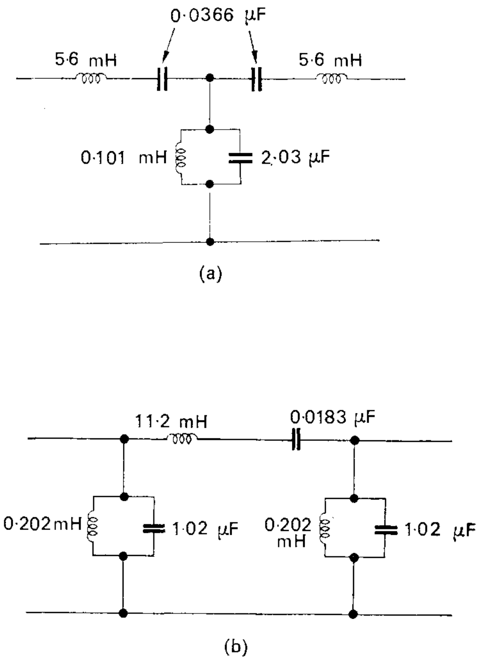


FIG. 1.25. 10–12·1 kHz band-pass filter.

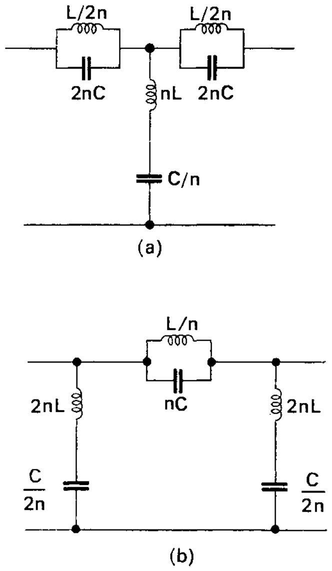


FIG. 1.26. Constant-*k* band-stop filter.

is to be stopped, but they must all be resonant at the same frequency. Hence the series inductance is decreased by a factor n and the series capacitance is increased by a factor n while the shunt inductance is increased and the shunt capacitance decreased by the same ratio. Then

$$Z_1 = \frac{j\omega(L/n)\times 1/(j\omega nC)}{j\omega(L/n)+1/(j\omega C)}$$

$$= \frac{j\omega L}{n(1-\omega^2 LC)}$$

$$= \frac{j\omega L}{n\left[1-\left(\frac{\omega}{\omega_0}\right)^2\right]},$$

where $$\omega_0^2 = \frac{1}{LC}$$

and $$Z_2 = j\omega nL+\frac{n}{j\omega C}$$

$$= j\omega nL\left[1-\frac{1}{\omega^2 LC}\right]$$

$$= j\omega nL\left[1-\left(\frac{\omega_0}{\omega}\right)^2\right].$$

Then, in the pass band,

$$\cos\beta = 1+\frac{Z_1}{2Z_2}$$

$$= 1+\frac{1}{2n^2\left[1-\left(\frac{\omega}{\omega_0}\right)^2\right]\left[1-\left(\frac{\omega_0}{\omega}\right)^2\right]},$$

i.e. $$\left[1-\left(\frac{\omega}{\omega_0}\right)^2\right]\left[1-\left(\frac{\omega_0}{\omega}\right)^2\right] = -\frac{1}{2n^2(1-\cos\beta)}.$$

If $\cos\beta = 1$, $\omega = 0$ or ∞ corresponding to extremes of the pass bands.

If $\cos\beta = -1$,

$$\left[1-\left(\frac{\omega}{\omega_0}\right)^2\right]\left[1-\left(\frac{\omega_0}{\omega}\right)^2\right] = -\frac{1}{4n^2},$$

i.e.
$$\left[1-\left(\frac{\omega}{\omega_0}\right)^2\right]^2 = \frac{1}{4n^2}\left(\frac{\omega}{\omega_0}\right)^2,$$

i.e.
$$1-\left(\frac{\omega}{\omega_0}\right)^2 = \pm\frac{1}{2n}\left(\frac{\omega}{\omega_0}\right)$$

or
$$\left(\frac{\omega}{\omega_0}\right)^2 \pm \frac{1}{2n}\left(\frac{\omega}{\omega_0}\right) - 1 = 0,$$

$$\left(\frac{\omega}{\omega_0}\right) = \pm\frac{1}{4n} \pm \sqrt{\left[\left(\frac{1}{4n}\right)^2+1\right]}.$$

The possible solutions are

$$\left(\frac{\omega}{\omega_0}\right) = \sqrt{\left[\left(\frac{1}{4n}\right)^2+1\right]} \pm \frac{1}{4n}.$$

Hence the other extremes of the pass bands are:

$$f_1 = f_0\left\{\sqrt{\left[\left(\frac{1}{4n}\right)^2+1\right]} + \frac{1}{4n}\right\},$$

$$f_2 = f_0\left\{\sqrt{\left[\left(\frac{1}{4n}\right)^2+1\right]} - \frac{1}{4n}\right\},$$

and the stop band covers a frequency range $f_0/2n$.

Within the pass band,

$$\cos\beta = 1+\frac{Z_1}{2Z_2}$$

$$= 1-\frac{\omega^2\omega_0^2}{2n^2(\omega^2-\omega_0^2)^2}$$

or
$$1-2\sin^2\frac{\beta}{2} = 1-2\left[\frac{\omega\omega_0}{2n(\omega^2-\omega_0^2)}\right]^2,$$

i.e.
$$\beta = 2\sin^{-1}\pm\left[\frac{\omega\omega_0}{2n(\omega^2-\omega_0^2)}\right].$$

As in the previous cases, this shows that the angle by which the output current lags or leads the input current follows an inverse sine relationship if the section is one of an infinite number of sections.

In the attenuation band, when $\beta = n\pi$, $\cos\beta = 1$,

i.e.
$$\pm\cosh\alpha = 1-\frac{\omega^2\omega_0^2}{2n^2(\omega^2-\omega_0^2)^2}.$$

Cosh α must be positive, and hence the negative sign must be taken. Then

$$2\cosh^2\frac{\alpha}{2}-1 = 2\left[\frac{\omega\omega_0}{2n(\omega^2-\omega_0^2)}\right]^2-1$$

or
$$\alpha = 2\cosh^{-1}\pm\left[\frac{\omega\omega_0}{2n(\omega^2-\omega_0^2)}\right].$$

This shows that within the attenuation band the attenuation follows an inverse cosh relationship, the positive and negative signs being chosen for frequencies between f_0 and f_1 and f_2 and f_0 respectively. The complete relationship is shown in Fig. 1.27.

Within the stop band when ω is less than ω_0 the negative sign must be chosen in the expression $\beta = 2\sin^{-1}\pm[\omega\omega_0/2n(\omega^2-\omega_0^2)]$

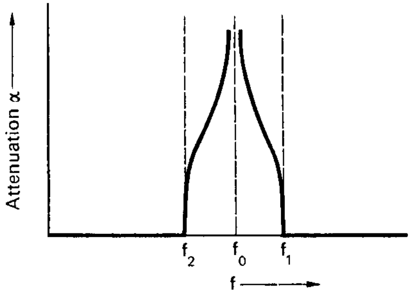


FIG. 1.27. Attenuation coefficient of a band-stop filter.

and by a similar reasoning to that employed in the low-pass filter the output current will lag the input current by π. If ω is greater than ω_0, still within the pass band, the output current will lead the input current by π. The complete variation in the phase of the output current with respect to the input current is shown in Fig. 1.28.

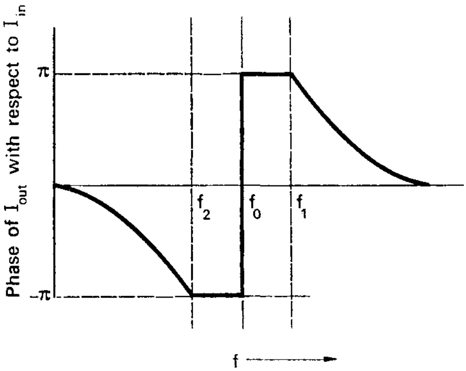


FIG. 1.28. Phase relation between input and output currents of a band-stop filter.

The characteristic impedance of the T-type band-stop section is given by

$$Z_{0T} = \sqrt{\left[\left(\frac{Z_1}{2}\right)^2 + Z_1 Z_2\right]},$$

where $Z_1 = \dfrac{j\omega L}{n\left[1-\left(\dfrac{\omega}{\omega_0}\right)^2\right]}$ and $Z_2 = j\omega n L\left[1-\left(\dfrac{\omega_0}{\omega}\right)^2\right]$.

Therefore
$$Z_{0T} = \sqrt{\left\{-\frac{\omega^2 L^2}{n^2\left[1-\left(\dfrac{\omega}{\omega_0}\right)^2\right]^2} + \omega_0^2 L^2\right\}}$$
$$= \sqrt{\left\{-\frac{\omega^2 L^2}{n^2\left[1-\left(\dfrac{\omega}{\omega_0}\right)^2\right]^2} + \frac{L}{C}\right\}},$$

where $$\omega_0^2 = \frac{1}{LC}.$$

At low frequencies Z_{0T} is $\sqrt{(L/C)}$ falling to zero at f_2. At high frequencies Z_{0T} is also $\sqrt{(L/C)}$ again falling to zero at f_1. Within the stop band, Z_{0T} is inductive rising to infinity at f_0, the geometric mean of the limits of the stop band. For the π section,

$$Z_{0\pi} = \frac{Z_1 Z_2}{Z_{0T}}$$

$$= \frac{\sqrt{(L/C)}}{\sqrt{\left\{1 - \dfrac{\omega^2}{n^2\omega_0^2\left[1-\left(\dfrac{\omega}{\omega_0}\right)^2\right]^2}\right\}}}.$$

This is also $\sqrt{(L/C)}$ at zero and infinite frequencies, rising to infinity at the limits of the stop band and becoming capacitive within the attenuation band. The complete variation in characteristic impedance of prototype band-stop filters is shown in Fig. 1.29.

As in previous cases, correct matching of the filter cannot be obtained over the complete frequency band. The geometric mean

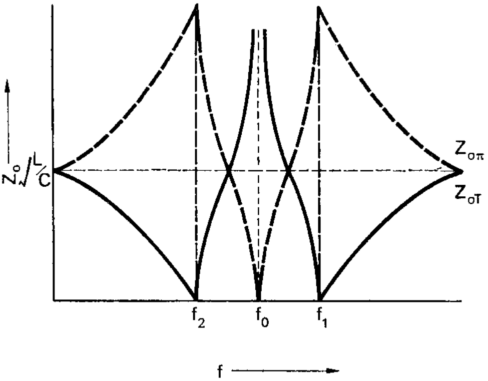


FIG. 1.29. Characteristic impedance of a band-stop filter.

of either pass band may be chosen, and in this case the design frequency is either ∞ or 0, giving $R_0 = \sqrt{(L/C)}$. Hence, since

$$f_0 = \sqrt{f_1 f_2} = \frac{1}{2\pi\sqrt{(LC)}},$$

$$L = \frac{R_0}{2\pi f_0}, \quad C = \frac{1}{2\pi f_0 R_0}.$$

The series and shunt elements may then be determined by putting

$$n = \frac{f_0}{2(f_1 - f_2)}.$$

If a prototype band-stop filter for the range 10 kHz to 12·1 kHz in a 75 Ω system is considered,

$$f_0 = \sqrt{(10 \times 12{\cdot}1)} = 11 \text{ kHz},$$

$$L = \frac{75}{2\pi \times 11 \times 10^3} = 1{\cdot}06 \text{ mH},$$

$$C = \frac{10^6}{2\pi \times 11 \times 10^3 \times 75} = 0{\cdot}193\ \mu\text{F},$$

$$n = \frac{11 \times 10^3}{2 \times 2{\cdot}1 \times 10^3} = 2{\cdot}62,$$

i.e. the series inductance is L/n = 0·405 mH;

the series capacitance is nC = 0·505 μF;

the shunt inductance is nL = 2·78 mH;

the shunt capacitance is C/n = 0·0736 μF.

Complete prototype T and π band-stop filters are shown in Fig. 1.30.

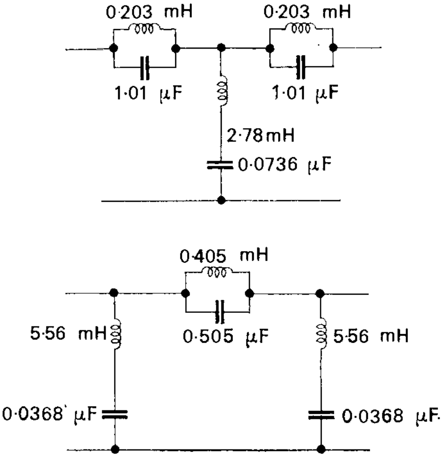


FIG. 1.30. 10–12·1 kHz band-stop filter.

1.4. *m*-Derived Sections

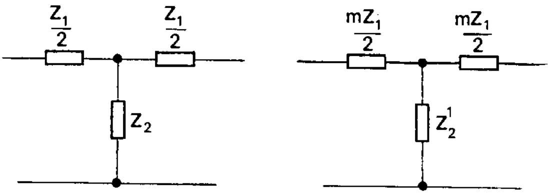
One of the main disadvantages of prototype filters is the low rate of rise of attenuation beyond cut-off. This may be counteracted by the use of *m*-derived sections. The analysis of such sections requires the use of the theorem that states that if two sections have the same characteristic impedance they have the same pass band.

An *m*-derived *T* section may be constructed by multiplying the component values in the series arm by a factor *m* and calculating the new value of the shunt components, i.e. from Fig. 1.31 (a) and (b), for the same characteristic impedance,

$$Z_{0T} = \sqrt{\left[\left(\frac{Z_1}{2}\right)^2 + Z_1Z_2\right]} = \sqrt{\left[\left(\frac{mZ_1}{2}\right)^2 + mZ_1Z_2^1\right]},$$

i.e.
$$Z_2^1 = \frac{(1-m^2)Z_1}{4m} + \frac{Z_2}{m}.$$

For this to be valid over the complete range of frequency (i.e.



(a) (b)

FIG. 1.31. Derivation of an *m*-derived *T* section.

without changing the nature of the reactances), *m* must be positive and less than unity,

i.e. $$0 < m < 1.$$

The *m*-derived section is then as shown in Fig. 1.32.

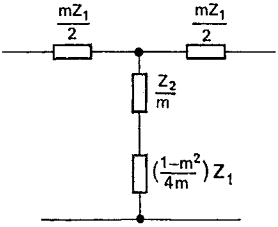
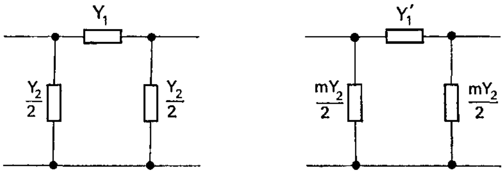


FIG. 1.32. *m*-Derived *T* section.

A similar treatment may be employed for the π section, but, as indicated previously, it is easier to work in admittances and multiply the shunt admittance by a factor *m*. Prototype and *m*-derived π sections are shown in Fig. 1.33 (a) and (b).



(a) (b)

FIG. 1.33. Derivation of an *m*-derived π section.

For the same characteristic admittance,

$$Y_{0\pi} = \sqrt{\left[\left(\frac{Y_2}{2}\right)^2 + Y_1Y_2\right]} = \sqrt{\left[\left(\frac{mY_2}{2}\right)^2 + mY_2Y_1'\right]}$$

i.e.
$$Y_1' = \frac{(1-m^2)Y_2}{4m} + \frac{Y_1}{m}.$$

This has the same restriction as the T section, i.e. $0 < m < 1$' and the m-derived π section is shown in its impedance form in Fig. 1.34.

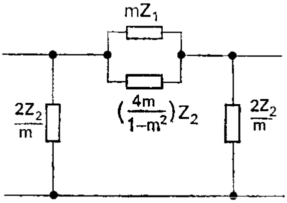
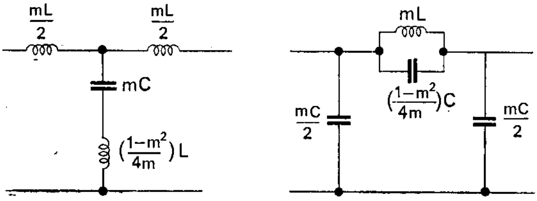


FIG. 1.34. m-Derived π section.

1.4.1. m-Derived low-pass filters

As stated previously, a prototype low-pass filter has a relatively low rate of rise of attenuation beyond cut-off. This rises to an infinite value at an infinite frequency. If the frequency at which infinite attenuation occurs is reduced, there will be a corresponding increase in the rate of rise of attenuation from cut-off at the cost of reduced attenuation at the higher frequencies.



(a) (b)

FIG. 1.35. m-Derived low-pass filter.

Infinite attenuation occurs in a T section at the series resonant frequency of the shunt arm, while for a π section infinite attenuation occurs at the parallel resonant frequency of the series arm. These sections are shown in Fig. 1.35 (a) and (b), and in each case the resonant angular frequency, i.e. the angular frequency at which there is infinite attenuation, is given by

$$\omega_\infty^2 = \frac{1}{mC\left(\dfrac{1-m^2}{4m}\right)L}.$$

$$= \frac{4}{LC(1-m^2)}$$

$$= \frac{\omega_c^2}{(1-m^2)}.$$

Typical attenuation curves for an m-derived low-pass filter are shown in Fig. 1.36.

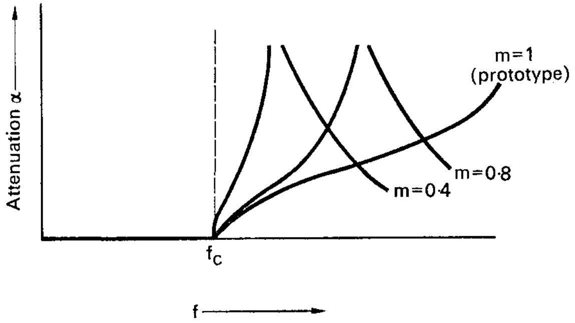


FIG. 1.36. Attenuation coefficient of an m-derived low-pass filter.

As an example of the use of m-derived low-pass filters, consider the previous example, i.e. a low-pass filter having a cut-off frequency of 10 kHz in a 75 Ω system with the additional limitation that a very high attenuation is required at 12 kHz.

Hence $$1{\cdot}44 \times 10^8 = \frac{10^8}{1-m^2}$$

or $$m^2 = 0{\cdot}306,$$

i.e. $$m = 0{\cdot}553.$$

Then the series inductance is $0{\cdot}553 \times 2{\cdot}39 = 1{\cdot}32$ mH;

the shunt capacitance is $0{\cdot}553 \times 0{\cdot}424 = 0{\cdot}234$ μF;

the additional shunt inductance is $\left(\frac{1-0{\cdot}306}{4 \times 0{\cdot}553}\right) \times 2{\cdot}39 =$ $0{\cdot}75$ mH;

the additional capacitance in the series arm for the π section is $\left(\frac{1-0{\cdot}306}{4 \times 0{\cdot}553}\right) \times 0{\cdot}424 = 0{\cdot}133$ μF.

Such T and π, m-derived low-pass filters are shown in Fig. 1.37. In general, to achieve a high rate of rise of attenuation, a somewhat lower value of m would be chosen, say $m = 0{\cdot}3$ or $0{\cdot}4$.

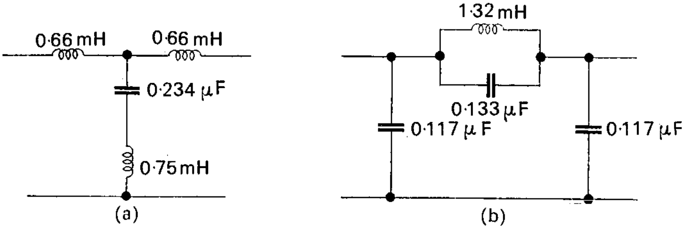


FIG. 1.37. 10 kHz m-derived low-pass filter.

1.4.2. *m-Derived high-pass filters*

A similar approach to that used with low-pass filters may be employed with high-pass filters, but in this case the infinite attenuation occurs at a chosen frequency between zero and cut off.

Typical m-derived, high-pass sections are shown in Fig. 1.38(a) and (b). Infinite attenuation occurs at the series resonant frequency

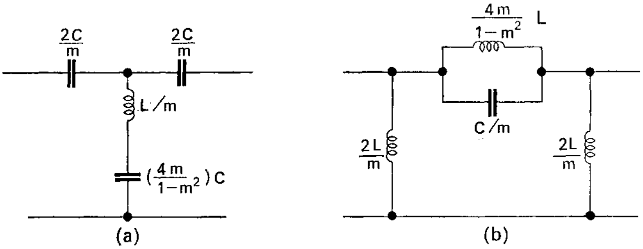


FIG. 1.38. *m*-Derived high-pass filter.

of the shunt arm for the *T* section and the parallel resonant frequency of the series arm for the π section.

In each case

$$\omega_\infty^2 = \frac{1}{\dfrac{L}{m}\left(\dfrac{4m}{1-m^2}\right)C}$$
$$= (1-m^2)\omega_c^2.$$

Typical attenuation curves for an *m*-derived high-pass filter are shown in Fig. 1.39.

As an example of an *m*-derived high-pass filter, consider the previous example in Fig. 1.20, i.e. a high-pass filter having a cut-off frequency of 10 kHz in a 75 Ω system with the additional limitation

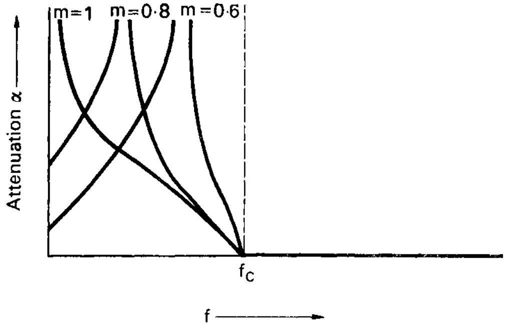


FIG. 1.39. Attenuation coefficient of an *m*-derived high-pass filter.

that a very high attenuation is required at 8 kHz,

i.e. $$64\times 10^6 = (1-m^2)10^8$$

or $$m^2 = 0{\cdot}36,$$

i.e. $$m = 0{\cdot}6.$$

Then the series capacitance is $\dfrac{0{\cdot}106}{0{\cdot}6} = 0{\cdot}177\ \mu\text{F}$;

the shunt inductance is $\dfrac{0{\cdot}597}{0{\cdot}6} = 0{\cdot}995\ \text{mH}$;

the additional shunt capacitance is $\dfrac{4\times 0{\cdot}6}{1-0{\cdot}36}\times 0{\cdot}106 =$ $0{\cdot}398\ \mu\text{F}$;

the additional inductance in the series arm for the π section is $\dfrac{4\times 0{\cdot}6}{1-0{\cdot}36}\times 0{\cdot}597 = 2{\cdot}24\ \text{mH}$.

Such T and π, m-derived high-pass filters are shown in Fig. 1.40. In general, m would be 0·3 or 0·4 to achieve a high rate of rise in attenuation from cut-off.

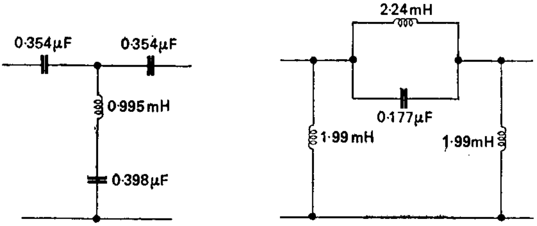
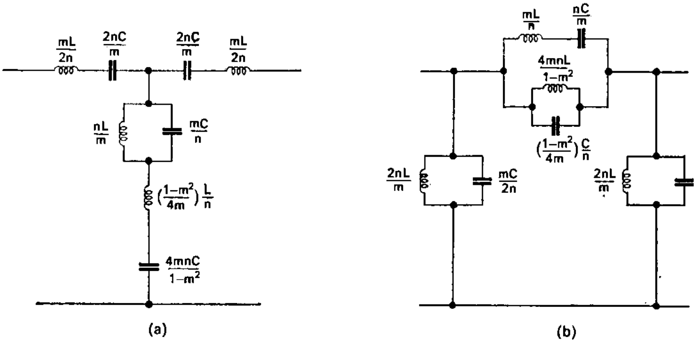


FIG. 1.40. 10 kHz m-derived high-pass filter.

1.4.3. m-Derived band-pass filters

Following the same pattern as previous m-derived sections, infinite attenuation may be obtained at specified frequencies outside the pass band. Typical m-derived band-pass sections are shown in Fig. 1.41 (a) and (b).



(a) (b)

FIG. 1.41. *m*-Derived band-pass filters.

Infinite attenuation occurs at the series resonant frequencies of the shunt arm for the *T* section and at the parallel resonant frequencies of the series arm of the π section. For the *T* section,

$$Z_{sh} = \frac{(1-m^2)}{j\omega 4mnC} + j\omega\left(\frac{1-m^2}{4m}\right)\frac{L}{n} + \frac{\dfrac{j\omega nL}{m}\times\dfrac{n}{j\omega mC}}{\dfrac{j\omega nL}{m}+\dfrac{n}{j\omega mC}}$$

$$= \left(\frac{1-m^2}{4m}\right)\frac{(1-\omega^2LC)}{j\omega nC} + \frac{j\omega nL}{m(1-\omega^2LC)};$$

putting $\omega_0^2 = \dfrac{1}{LC}$,

$$Z_{sh} = \frac{(1-m^2)\left[1-\left(\dfrac{\omega}{\omega_0}\right)^2\right]^2 - 4n^2\left(\dfrac{\omega}{\omega_0}\right)^2}{j\omega 4mnC\left[1-\left(\dfrac{\omega}{\omega_0}\right)^2\right]}.$$

For this to be zero,

$$(1-m^2)\left[1-\left(\frac{\omega_\infty}{\omega_0}\right)^2\right]^2 - 4n^2\left(\frac{\omega_\infty}{\omega_0}\right)^2 = 0$$

or $$\left[1-\left(\frac{\omega_\infty}{\omega_0}\right)^2\right] = \pm\frac{2n}{\sqrt{(1-m^2)}}\times\left(\frac{\omega_\infty}{\omega_0}\right),$$

i.e. $$\left(\frac{\omega_\infty}{\omega_0}\right)^2 \pm \frac{2n}{\sqrt{(1-m^2)}}\left(\frac{\omega_\infty}{\omega_0}\right)-1 = 0,$$

i.e. $$\left(\frac{\omega_\infty}{\omega_0}\right) = \pm\frac{n}{\sqrt{(1-m^2)}} \pm \sqrt{\left[\frac{n^2}{(1-m^2)}+1\right]}.$$

The possible solutions of this equation are

$$\left(\frac{\omega_\infty}{\omega_0}\right) = \sqrt{\left[\frac{n^2}{(1-m^2)}+1\right]} \pm \frac{n}{\sqrt{(1-m^2)}}.$$

The series admittance for the π section is

$$Y_{se} = \frac{(1-m^2)}{j\omega 4mnL} + j\omega\left(\frac{1-m^2}{4m}\right)\frac{C}{n} + \frac{\dfrac{j\omega nC}{m}\times\dfrac{n}{j\omega mL}}{\dfrac{j\omega nC}{m}+\dfrac{n}{j\omega mL}}.$$

This is comparable with the shunt impedance of the T section and equating to zero yields the same result for the frequencies at which infinite attenuation occurs.

Typical attenuation curves for m-derived band-pass filters are shown in Fig. 1.42.

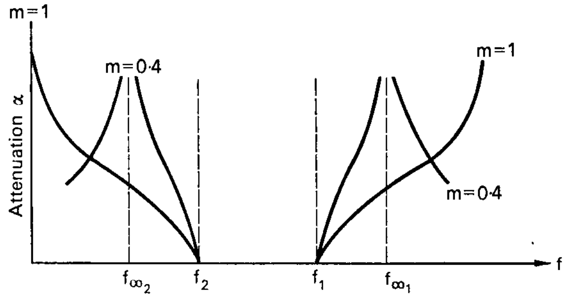


FIG. 1.42. Attenuation coefficient of an m-derived band-pass filter.

As an example of this type, the band-pass filter previously discussed in section 1.3.3 may be considered, i.e. a band-pass filter for the range 10–12·1 kHz in a 75 Ω system. If there is an additional limitation of a very high attenuation at 9 kHz in the lower attenuation band,

$$\frac{9}{11} = \sqrt{\left[\frac{0{\cdot}095^2}{(1-m^2)}+1\right]} - \frac{0{\cdot}095}{\sqrt{(1-m^2)}},$$

i.e.
$$0{\cdot}818+\frac{0{\cdot}095}{\sqrt{(1-m^2)}} = \sqrt{\left[\frac{0{\cdot}095^2}{(1-m^2)}+1\right]},$$

$$0{\cdot}67+\frac{0{\cdot}095^2}{(1-m^2)}+\frac{0{\cdot}155}{\sqrt{(1-m^2)}} = \frac{0{\cdot}095^2}{(1-m^2)}+1,$$

i.e.
$$1-m^2 = \left(\frac{0{\cdot}155}{1-0{\cdot}67}\right)^2$$

or
$$m^2 = 0{\cdot}78,$$
$$m = 0{\cdot}883.$$

This also gives a very high attenuation at 13·44 kHz in the upper attenuation band.

Then the series inductance is $0{\cdot}883\times 11{\cdot}2 = 9{\cdot}9$ mH;

the series capacitance is $\dfrac{0{\cdot}0183}{0{\cdot}883} = 0{\cdot}0207\ \mu\text{F}$;

the shunt inductance is $\dfrac{0{\cdot}101}{0{\cdot}883} = 0{\cdot}114$ mH;

the shunt capacitance is $0{\cdot}883\times 2{\cdot}03 = 1{\cdot}79\ \mu\text{F}$;

the additional series inductance for the shunt arm of the T section is $\left(\dfrac{1-0{\cdot}78}{4\times 0{\cdot}883}\right)\times 11{\cdot}2 = 0{\cdot}70$ mH;

the additional series capacitance for the shunt arm of the T section is $\dfrac{4\times 0{\cdot}883}{1-0{\cdot}78}\times 0{\cdot}0183 = 0{\cdot}294\ \mu\text{F}$;

the additional shunt inductance for the series arm of the π section is $\frac{4\times0{\cdot}883}{(1-0{\cdot}78)}\times0{\cdot}101 \quad = 1{\cdot}62$ mH;

the additional shunt capacitance for the series arm of the π section is $\frac{(1-0{\cdot}78)}{4\times0{\cdot}883}\times2{\cdot}03 \quad = 0{\cdot}126\ \mu$F.

Such T and π, band-pass, m-derived filters are shown in Fig. 1.43.

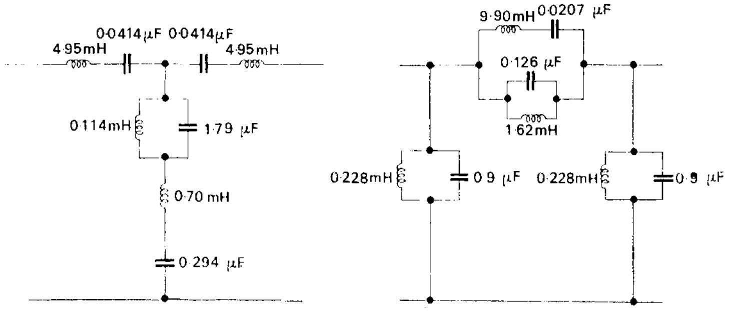
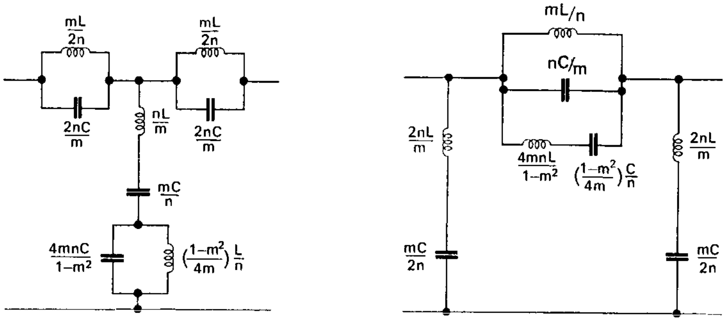


FIG. 1.43. 10–12·1 kHz m-derived band-pass filter.

1.4.4. m-Derived band-stop filters

Typical m-derived band-stop sections are illustrated in Fig. 1.44 (a) and (b).



(a) (b)

FIG. 1.44. m-Derived band-stop filters.

The frequencies of infinite attenuation occur at the series-resonant frequencies of the shunt arm for the T section and the parallel-resonant frequencies for the π section. For the T section,

$$Z_{sh} = \frac{j\omega nL}{m} + \frac{n}{j\omega mC} + \frac{j\omega\left(\frac{1-m^2}{4m}\right)\frac{L}{n}\times\frac{(1-m^2)}{j\omega 4mnC}}{j\omega\left(\frac{1-m^2}{4m}\right)\frac{L}{n}+\frac{(1-m^2)}{j\omega 4mnC}}$$

$$= \frac{n}{j\omega mC}(1-\omega^2 LC) + \frac{j\omega\left(\frac{1-m^2}{4m}\right)\frac{L}{n}}{1-\omega^2 LC};$$

putting $\omega_0^2 = \dfrac{1}{LC}$,

$$Z_{sh} = \frac{n}{j\omega mC}\left(1-\frac{\omega^2}{\omega_0^2}\right) + \frac{j\omega\left(\frac{1-m^2}{4m}\right)\frac{L}{n}}{\left(1-\frac{\omega^2}{\omega_0^2}\right)}$$

$$= \frac{4n^2\left[1-\left(\frac{\omega}{\omega_0}\right)^2\right]^2-(1-m^2)\left(\frac{\omega}{\omega_0}\right)^2}{4j\omega mnC\left[1-\left(\frac{\omega}{\omega_0}\right)^2\right]}.$$

For this to be zero,

$$4n^2\left[1-\left(\frac{\omega_\infty}{\omega_0}\right)^2\right]^2-(1-m^2)\left(\frac{\omega_\infty}{\omega_0}\right)^2 = 0,$$

$$1-\left(\frac{\omega_\infty}{\omega_0}\right)^2 = \pm\frac{\sqrt{(1-m^2)}}{2n}\left(\frac{\omega_\infty}{\omega_0}\right),$$

i.e.

$$\left(\frac{\omega_\infty}{\omega_0}\right)^2 \pm \frac{\sqrt{(1-m^2)}}{2n}\left(\frac{\omega_\infty}{\omega_0}\right) - 1 = 0$$

or

$$\left(\frac{\omega_\infty}{\omega_0}\right) = \pm\frac{\sqrt{(1-m^2)}}{4n} \pm \sqrt{\left[\frac{(1-m^2)}{16n^2}+1\right]}.$$

The possible solutions of this equation are

$$\left(\frac{\omega_\infty}{\omega_0}\right) = \sqrt{\left[\frac{(1-m^2)}{16n^2}+1\right]} \pm \frac{\sqrt{(1-m^2)}}{4n}.$$

The series admittance for the π section is

$$Y_{se} = \frac{n}{j\omega mL} + \frac{j\omega nC}{m} + \frac{\dfrac{(1-m^2)}{j\omega 4mnL} \times \dfrac{j\omega C(1-m^2)}{4mn}}{\dfrac{(1-m^2)}{j\omega 4mnL} + \dfrac{j\omega C(1-m^2)}{4mn}}.$$

This is comparable with the shunt impedance of the *T* section, and equating to zero yields the same result for the frequencies at which infinite attenuation occurs.

Typical attenuation curves for *m*-derived band-stop filters are shown in Fig. 1.45.

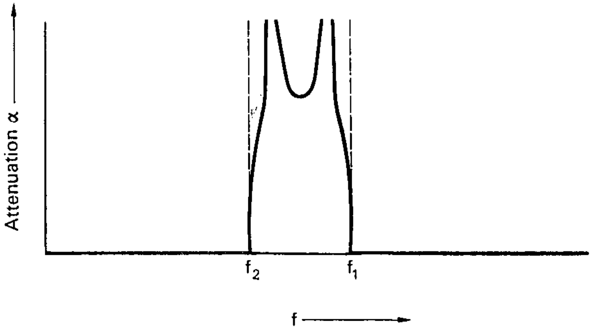


FIG. 1.45. Attenuation coefficient of an *m*-derived band-stop filter.

As an example of this type, the band-stop filter previously discussed in section 1.3.4 may be considered, i.e. a band-stop filter for the range 10–12·1 kHz in a 75 Ω system. If there is an additional limitation of a very high attenuation at 10·45 kHz,

$$\left(\frac{10{\cdot}45}{11}\right) = \sqrt{\left[\frac{(1-m^2)}{16\times 2{\cdot}62^2}+1\right]} - \frac{\sqrt{(1-m^2)}}{4\times 2{\cdot}62},$$

i.e.
$$0{\cdot}95+\frac{\sqrt{(1-m^2)}}{4\times 2{\cdot}62}=\sqrt{\left[\frac{(1-m^2)}{16\times 2{\cdot}62^2}+1\right]},$$

$$0{\cdot}903+\frac{(1-m^2)}{16\times 2{\cdot}62^2}+\frac{1{\cdot}9\sqrt{(1-m^2)}}{10{\cdot}48}=\frac{(1-m^2)}{16\times 2{\cdot}62^2}+1,$$

i.e.
$$1-m^2=\left(\frac{0{\cdot}097\times 10{\cdot}48}{1{\cdot}9}\right)^2$$

or
$$m^2=0{\cdot}714,$$

$$m=0{\cdot}845.$$

Then the inductance in the series arm is $0{\cdot}845\times 0{\cdot}405=0{\cdot}342$ mH;

the capacitance in the series arm is $\frac{0{\cdot}505}{0{\cdot}845}$ = 0·6 μF;

the shunt inductance is $\frac{2{\cdot}78}{0{\cdot}845}$ = 3·29 mH;

the shunt capacitance is $0{\cdot}845\times 0{\cdot}0736$ = 0·062 μF;

the additional capacitance in the shunt arm for the *T* section is $\frac{4\times 0{\cdot}845\times 0{\cdot}505}{0{\cdot}286}$ = 5·98 μF;

the additional inductance in the shunt arm for the *T* section is $\frac{0{\cdot}286}{4\times 0{\cdot}845}\times 0{\cdot}405$ =0·0343 mH;

the additional series inductance for the π section is $\frac{4\times 0{\cdot}845}{0{\cdot}286}\times 2{\cdot}78$ = 32·9 mH;

the additional series capacitance for the π section is $\frac{0{\cdot}286}{4\times 0{\cdot}845}\times 0{\cdot}0736$ =0·0062 μF.

Such T and π, m-derived band-stop filters are shown in Fig. 1.46.

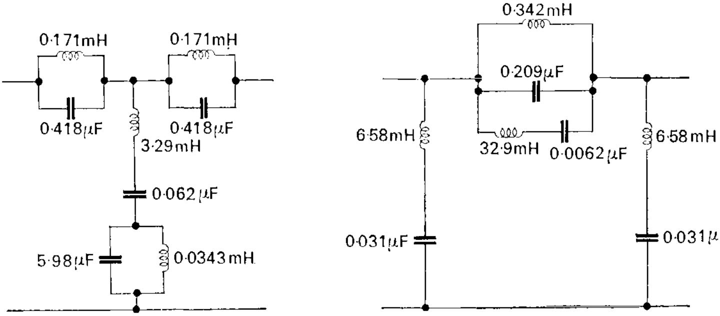


FIG. 1.46. 10–12·1 kHz m-derived band-stop filter.

1.5. m-Derived Half Section

One disadvantage with prototype sections, i.e. the low rate of rise of attenuation beyond cut-off, has already been discussed. The other main disadvantage is the variation of characteristic impedance over the pass band. This may be counteracted by using m-derived half sections for matching purposes.

For such asymmetrical sections, one cannot work in terms of characteristic impedance. Instead the image impedances must be considered, these being defined as two impedances such that if one is connected across two terminals the other is presented across the other two.

Figure 1.47 shows an m-derived T section split in two to produce

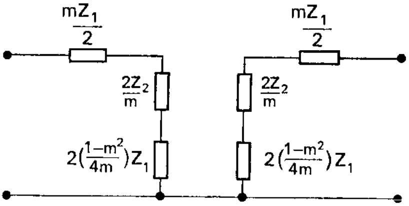

$\frac{mZ_1}{2}$
$\frac{mZ_1}{2}$
$\frac{2Z_2}{m}$
$\frac{2Z_2}{m}$
$2\left(\frac{1-m^2}{4m}\right)Z_1$
$2\left(\frac{1-m^2}{4m}\right)Z_1$

FIG. 1.47. Derivation of m-derived half T section.

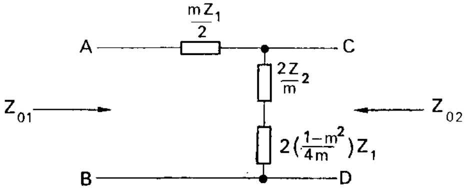


FIG. 1.48. *m*-derived half *T* section.

the half sections, and Fig. 1.48 the half section with appropriate image impedances.

Then

$$Z_{01} = \sqrt{Z_{oc1}Z_{sc1}}$$

where Z_{oc1} and Z_{sc1} are the open and short circuit impedances looking into terminals *AB*,

i.e.
$$\begin{aligned} Z_{01} &= \sqrt{\left[\frac{mZ_1}{2}+\frac{2Z_2}{m}+2\left(\frac{1-m^2}{4m}\right)Z_1\right]\frac{mZ_1}{2}} \\ &= \sqrt{\left[\frac{2Z_2}{m}+\frac{Z_1}{2m}\right]\frac{mZ_1}{2}} \\ &= \sqrt{\left[Z_1Z_2+\frac{Z_1^2}{4}\right]} \\ &= Z_{0T} \end{aligned}$$

and
$$Z_{02} = \sqrt{Z_{oc2}Z_{sc2}},$$

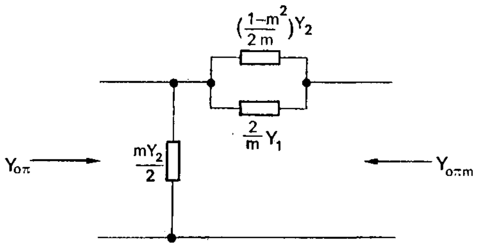


FIG. 1.49. *m*-Derived half *π* section.

where Z_{oc2} and Z_{sc2} are the open and short circuit impedances looking into terminals *CD*,

$$\text{i.e. } Z_{02}=\sqrt{\left[\frac{2Z_2}{m}+2\left(\frac{1-m^2}{4m}\right)Z_1\right]\left[\frac{\left[\frac{2Z_2}{m}+2\left(\frac{1-m^2}{4m}\right)Z_1\right]\frac{mZ_1}{2}}{\frac{mZ_1}{2}+\frac{2Z_2}{m}+2\left(\frac{1-m^2}{4m}\right)Z_1}\right]}$$

$$=\frac{\left[\frac{2Z_2}{m}+2\left(\frac{1-m^2}{4m}\right)Z_1\right]\frac{mZ_1}{2}}{\sqrt{\left\{\frac{mZ_1}{2}\left[\frac{mZ_1}{2}+\frac{2Z_2}{m}+2\left(\frac{1-m^2}{4m}\right)Z_1\right]\right\}}}$$

$$=\frac{\left(Z_1Z_2+\frac{Z_1^2}{4}-\frac{m^2Z_1^2}{4}\right)}{Z_{0T}}$$

$$=Z_{0T}-\frac{m^2Z_1^2}{4Z_{0T}}.$$

This is known as Z_{0Tm}.

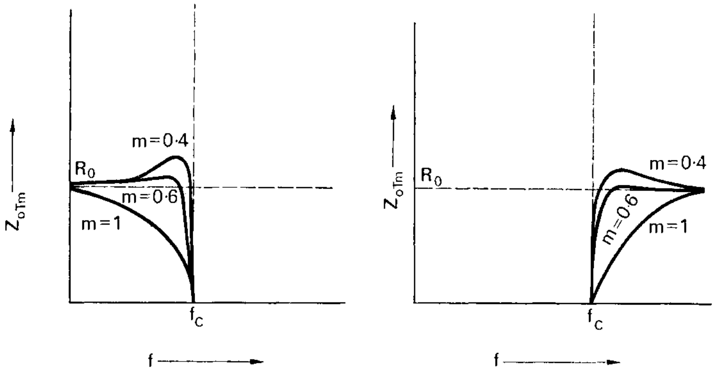
In a similar manner, the *m*-derived π section shown in Fig. 1.34 may be split in two as shown in Fig. 1.49.

Then

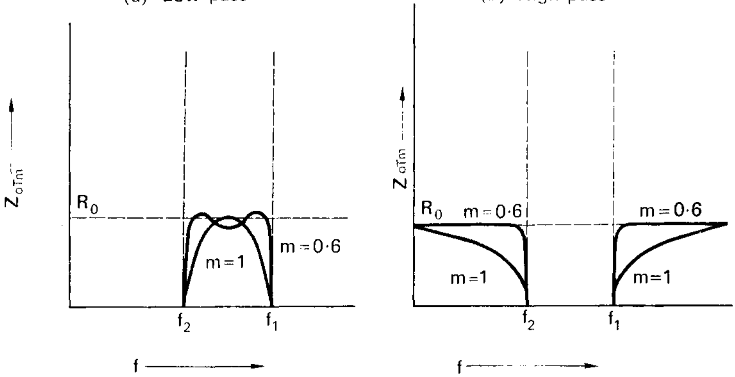
$$Y_{0\pi m}=Y_{0\pi}-\frac{m^2Y_2^2}{4Y_{0\pi}}.$$

Using half sections, the image impedance Z_{0T} enables correct matching to full filter sections to be achieved while the other image impedance Z_{0Tm} may be matched to the system. The value of Z_{0Tm} (or $Y_{0\pi m}$) obviously depends on *m* and typical variations of Z_{0Tm} are shown in Fig. 1.50 (a) to (d).

Obviously, from the curves shown in Fig. 1.50, the optimum value of *m* for the matching section is 0·6. Using this value, matching sections for the cases previously discussed may be calculated.



(a) Low pass (b) High pass



(c) Band pass (d) Band stop

FIG. 1.50. Characteristic impedance of *m*-derived half sections.

1.5.1. Matching section for low-pass filter

If the filter previously discussed in section 1.4.1 is considered, the matching half section for a low-pass filter with a cut-off frequency of 10 kHz for a 75 Ω system may be derived by putting $m = 0{\cdot}6$.

Then the series inductance is $0{\cdot}6 \times 2{\cdot}39 = 1{\cdot}43$ mH;

the shunt capacitance is $0{\cdot}6 \times 0{\cdot}424 = 0{\cdot}254$ μF;

the additional shunt inductance for the T section is

$$\frac{(1-0{\cdot}36)}{4\times 0{\cdot}6}\times 2{\cdot}39 \qquad = 0{\cdot}67 \text{ mH};$$

the additional capacitance in the series arm of the π section is

$$\frac{(1-0{\cdot}36)}{(4\times 0{\cdot}6)}\times 0{\cdot}424 \qquad = 0{\cdot}113\ \mu\text{F}.$$

The appropriate matching sections for the low-pass filters are then as shown in Fig. 1.51.

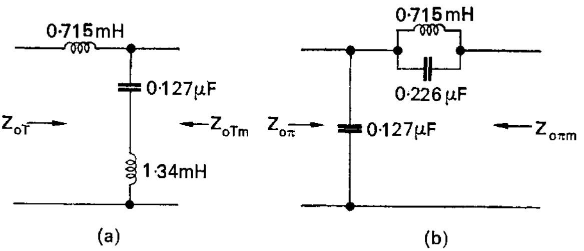


FIG. 1.51. *m*-Derived low-pass half sections.

1.5.2. Matching section for a high-pass filter

This, again, may be derived from the discussion in section 1.4.2, i.e. a high-pass filter with a cut-off frequency of 10 kHz in a 75 Ω system. In fact the section discussed was evaluated for $m = 0{\cdot}6$ and the matching half sections are as shown in Fig. 1.52 (a) and (b).

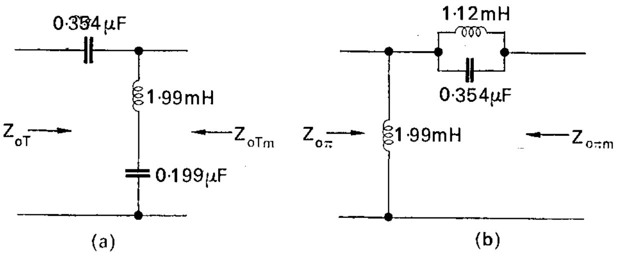


FIG. 1.52. *m*-Derived high-pass half sections.

1.5.3. Matching section for a band-pass filter

If *m* is put equal to 0·6 in the filter derived in section 1.4.3, matching half sections for a band-pass filter may be obtained.

Then the series inductance is $0{\cdot}6\times 11{\cdot}2 = 6{\cdot}72$ mH;

the series capacitance is $\dfrac{0{\cdot}0183}{0{\cdot}6} = 0{\cdot}0305\ \mu\text{F}$;

the shunt inductance is $\dfrac{0{\cdot}101}{0{\cdot}6} = 0{\cdot}168$ mH;

the shunt capacitance is $0{\cdot}6\times 2{\cdot}03 = 1{\cdot}22\ \mu\text{F}$;

the additional shunt inductance for the *T* section is

$$\left(\frac{1-0{\cdot}36}{4\times 0{\cdot}6}\right)\times 11{\cdot}2 = 2{\cdot}99\ \text{mH};$$

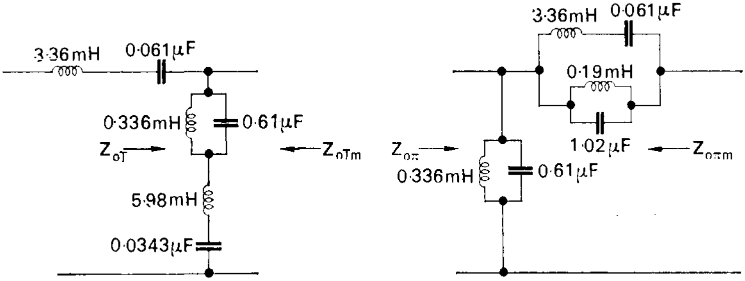
the additional shunt capacitance for the *T* section is

$$\left(\frac{4\times 0{\cdot}6}{1-0{\cdot}36}\right)\times 0{\cdot}0183 = 0{\cdot}0686\ \mu\text{F};$$

the additional inductance in the series arm of the π section is $\left(\dfrac{4\times 0{\cdot}6}{1-0{\cdot}36}\right)\times 0{\cdot}101 = 0{\cdot}379$ mH;

the additional capacitance in the series arm of the π section is $\left(\dfrac{1-0{\cdot}36}{4\times 0{\cdot}6}\right)\times 2{\cdot}03 = 0{\cdot}508\ \mu\text{F}$.

Using these values, the terminating half section for a band-pass filter for the range 10–12·1 kHz in a 75 Ω system are as shown in Fig. 1.53 (a) and (b). In practice, the shunt arm of the *T* half section is sometimes converted into an equivalent network comprising two series resonant circuits in parallel, while the series arm of the π half section is replaced by an equivalent network comprising two parallel resonant circuits in series.



(a) (b)

FIG. 1.53. *m*-Derived band-pass half sections.

1.5.4. Matching half section for a band-stop filter

As with previous matching sections, this simply involves putting $m = 0{\cdot}6$ in the *m*-derived discussion in section 1.4.4 and splitting the section into two. For the complete section:

the inductance in series arm is $0{\cdot}6 \times 0{\cdot}405 = 0{\cdot}243$ mH;

the capacitance in the series arm is $\dfrac{0{\cdot}505}{0{\cdot}6} = 0{\cdot}84$ μF;

the shunt inductance is $\dfrac{2{\cdot}78}{0{\cdot}6} = 4{\cdot}63$ mH;

the shunt capacitance is $0{\cdot}6 \times 0{\cdot}0736 = 0{\cdot}0442$ μF;

the additional capacitance in the shunt arm for the *T* section is $\dfrac{4 \times 0{\cdot}6}{0{\cdot}64} \times 0{\cdot}505 = 1{\cdot}89$ μF;

the additional inductance in the shunt arm of the *T* section is $\dfrac{0{\cdot}64}{4 \times 0{\cdot}6} \times 0{\cdot}405 = 0{\cdot}108$ mH;

the additional series inductance for the π section is

$\dfrac{4 \times 0{\cdot}6}{0{\cdot}64} \times 2{\cdot}78 = 10{\cdot}4$ mH;

the additional series capacitance for the π section is

$$\frac{0\cdot64}{4\times0\cdot6}\times0\cdot0736 \qquad = 0\cdot0196\ \mu\text{F}.$$

The matching T and π half sections are then as shown in Fig. 1.54. As in the case of the band pass filter the shunt arm of the T half section is sometimes converted into an equivalent network comprising two series resonant circuits in parallel and the series arm of the π half section replaced by two parallel resonant circuits in series.

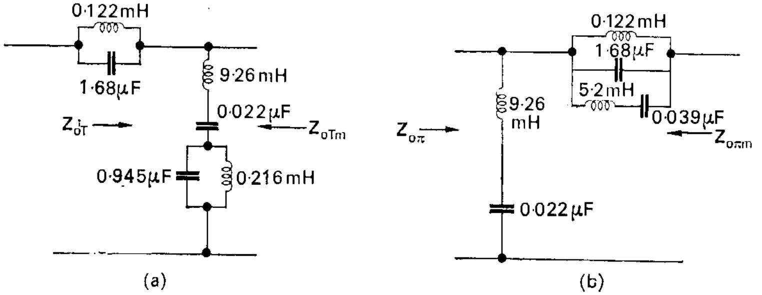


FIG. 1.54. *m*-Derived band-stop half sections.

1.6. Complete Filter Networks

As indicated in previous sections, a complete filter usually comprises, before reduction to its final form, several sections, i.e.

(1) A prototype section to maintain a high attenuation away from the pass band.

(2) An *m*-derived section to achieve a high rate of rise of attenuation from cut off.

(3) Matching sections to achieve a reasonable degree of matching over the majority of the pass band.

Any of the types previously discussed can now be combined to give a complete filter. If, for example, a complete low-pass filter with a cut-off frequency of 10 kHz for a 75 Ω system is required with high attenuation at 12 kHz and reasonable matching over the pass band, a suitable circuit would be as shown in Fig. 1.55. This reduces to Fig. 1.56.

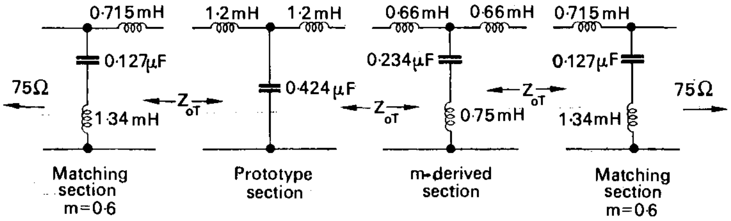


FIG. 1.55. Derivation of complete low-pass filter.

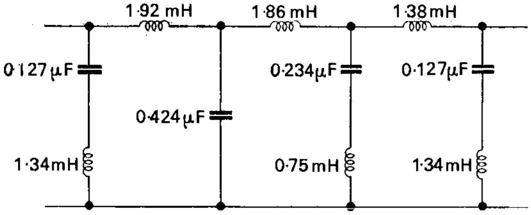


FIG. 1.56. Complete low-pass filter.

Questions on Chapter 1

Q.1. An unchanged capacitor C is connected in series with a resistance R and a d.c. voltage V applied to the combination. Derive an expression for the charge on the capacitor after a time t.

Q.2. A voltage $V_m \sin \omega t$ is connected to an unfluxed inductance L in series with a resistance R at the instant when $t = 0$. Show that the current flowing after a time t is given by

$$i = \frac{V_m}{\sqrt{(R^2+\omega^2L^2)}}\,[\varepsilon^{-Rt/L}\sin\theta + \sin(\omega t-\theta)],$$

where $$\tan\theta = \frac{\omega L}{R}.$$

Q.3. Explain, with reference to a low-pass filter, how the characteristic impedance of a symmetrical *T* section varies with frequency.

Q.4. Design a high-pass prototype filter with a cut-off angular frequency of 20,000 rad/s in a 500 Ω system. Discuss the limitations of such a filter and explain how these may be overcome.

Q.5. A high-pass filter with a cut-off frequency of 2 MHz is required for a 150 Ω feeder. Design such a filter including provision for high attenuation at 800 kHz and reasonable matching over the pass band.

Q.6. Design a prototype band-pass filter having cut-off frequencies of 10 kHz and 40 kHz in a 600 Ω system.

CHAPTER 2

Basic Physical Theory

IN THIS chapter the qualitative explanation of the physical properties of solids discussed in previous volumes will be extended and a more quantitative approach developed. The application to *pn* junctions and practical devices will be considered in the next chapter.

2.1. Electronic Structure of Matter

All atoms may be regarded as being composed of electrons rotating in circular or elliptical orbits about the nucleus of protons and neutrons. The simplest case is the hydrogen atom with one electron rotating in a circular orbit about a proton.

The radius of this orbit may be determined by using Bohr's basic assumption regarding any electron orbit. This assumption is that the angular momentum in any stable orbit must be an integral number of $h/2\pi$, where h is Planck's constant.

Thus if the moment of inertia of an electron is I and its angular velocity is ω,

$$I\omega = \frac{nh}{2\pi},$$

where n is an integer,

i.e.
$$m_e r^2 \frac{v}{r} = \frac{nh}{2\pi},$$

where m_e is the mass of an electron in kg, r is its distance from the nucleus in metres, and v its velocity in metres per second. Hence,

$$m_e r v = \frac{nh}{2\pi}. \tag{2.1.1}$$

Also equating the centrifugal force to the attractive force between electron and nucleus,

$$\frac{m_e v^2}{r} = \frac{e^2}{4\pi\varepsilon_0 r^2}, \tag{2.1.2}$$

where e is the charge on an electron and ε_0 is the permittivity of free space

$$\left(= \frac{1}{36\pi \times 10^9}\ \mathrm{F/m}\right).$$

If v is eliminated between eqns. (2.1.1) and (2.1.2),

$$r = \frac{n^2 h^2 \varepsilon_0}{\pi e^2 m_e}.$$

This equation enables the radii of possible orbits to be determined. Putting $n = 1$ and substituting for h, ε_0, π, and m, gives $r = 0{\cdot}529$ Å. This is the radius of the stable orbit nearest the nucleus and is the one normally occupied. Putting $n = 2$ gives a value for r of 2·11 Å, and is the radius of the second possible orbit. If, by some means, energy is supplied to raise the electron to the second possible orbit, it will eventually fall back to the inner stable orbit giving up its energy as radiation, the frequency f of which is given by

$$E_2 - E_1 = hf,$$

where E_2 and E_1 are the energies of the two orbits.

The radiated energy, in the form of visible light, is then characteristic of hydrogen.

The values of energy in these possible orbits may be obtained by adding the values of potential and kinetic energy (PE and KE), the former being the work done by an electron moving from infinity to a radius r,

$$\text{i.e.} \qquad \text{Total energy} = \text{PE}+\text{KE}$$

$$= \int_{\infty}^{r} \frac{e^2\,dr}{4\pi f_0 r^2} + \frac{m_e v^2}{2}$$

$$= -\frac{e^2}{4\pi\varepsilon_0 r} + \frac{e^2}{8\pi\varepsilon_0 r}$$

$$= -\frac{e^2}{8\pi\varepsilon_0 r} = -\frac{e^4 m_e}{8\varepsilon_0^2 n^2 h^2}\,.$$

Thus for the first orbit of hydrogen, i.e. $n = 1$, the total energy is -217×10^{-20} J or $-13{\cdot}7$ eV. The negative sign indicates that work must be done to remove the electron from the atom. The constant n is known as the principal quantum number and designates the K shell. If n is put equal to 2, the resulting energy value corresponds to that of the second shell, i.e. the L shell.

2.1.1. Quantum numbers

The previous paragraph has introduced the idea of the principal quantum number. In fact, there are four quantum numbers associated with each electron orbit.

(a) The principal quantum number (n)

For a single atom the electrons are arranged in shells of definite energy level where the electrostatic force of attraction is balanced by the centrifugal force on the electron. The principal quantum number n, which has integer values, gives the energy levels in these shells in arbitrary units. The value of these units depends on the element since the heavier elements with a greater positive charge on the nucleus give rise to a greater attractive force. That is, $n = 1$ corresponds to the K shell, $n = 2$ the L shell, etc. Thus a shell contains all the electrons with the same principal quantum number.

(b) The subsidiary quantum number (l)

This is a measure of the eccentricity of the electron orbits and has a slight effect on the electron energy. It may have values of $0 < l < n-1$ and corresponds to the subshells. That is, $n = 1$ can only have one value 0, i.e. one subshell: $n = 2$ can have values 0 or 1, i.e. two subshells. Thus a subshell contains all the electrons with the same principal and subsidiary quantum numbers.

(c) The magnetic quantum number (m)

This is associated with the rotation of the electrons about the nucleus and has a slight effect on the energy and may have values:

$$-l < m < l,$$

i.e.

$$\begin{aligned} l &= 0 \quad m = 0, \\ l &= 1 \quad m = -1, \ 0 \ \text{or} \ 1, \\ l &= 2 \quad m = -2, -1, 0, \ 1 \ \text{or} \ 2. \end{aligned}$$

(d) The spin quantum number (m_s)

This is associated with the rotation of the electron about its axis and may have two values $\pm\frac{1}{2}$.

The application of Pauli's exclusion principle, which states that no two electrons may have the same four quantum numbers, will now give the number of possible electron orbits:

$n = 1$, K shell, $l = 0$, 1s subshell, $m = 0$
$m_s = \pm\frac{1}{2}$, 2 electrons.

$n = 2$, L shell, $l = 0$, $2s$ subshell, $m = 0$,
$m_s = \pm\frac{1}{2}$, 2 electrons;

$l = 1$, $2p$ subshell, $m = -1, 0, 1$,
$m_s = \pm\frac{1}{2}$, 6 electrons.

$n = 3$, M shell, $l = 0$, $3s$ subshell, $m = 0$
$m_s = \pm\frac{1}{2}$, 2 electrons;

$l = 1$, $3p$ subshell, $m = -1, 0, 1$
$m_s = \pm\frac{1}{2}$, 6 electrons;

$l = 2$, $3d$ subshell, $m = -2, -1, 0, 1, 2$
$m_s = \pm\frac{1}{2}$, 10 electrons.

$n = 4$, N shell, $l = 0$, $4s$ subshell, $m = 0$,
$m_s = \pm\frac{1}{2}$, 2 electrons;

$l = 1$, $4p$ subshell, $m = -1, 0, 1$
$m_s = \pm\frac{1}{2}$, 6 electrons;

$l = 2$, $4d$ subshell, $m = -2, -1, 0, 1, 2$
$m_s = \pm\frac{1}{2}$, 10 electrons;

$l = 3$, $4f$ subshell, $m = -3, -2, -1, 0, 1, 2, 3$,
$m_s = \pm\frac{1}{2}$, 14 electrons.

This shows that the K shell has 2 possible electron states, the L shell 8, the M shell 18, and the N shell 32. The number of electron states in any shell is twice the square of the principal quantum number.

The significance of these quantum numbers lies in their application to the band theory of solids and, in particular, to semiconductor theory. The electrons in any shell only have the same energy when a single atom is considered. As soon as a collection of atoms is considered, the permissible energy levels spread out into bands. Each of these bands correspond to a subshell. The interaction between atoms and electrons is such that the electrons in certain subshells, e.g. in the fourth shell, may be more tightly bound to the nucleus than electrons in lower shells, e.g. the third shell.

A typical band structure for a solid is illustrated by Fig. 2.1. This shows the outer conduction band, the inner valence band, and dotted curves representing the energy binding electrons to the nuclei and the energy barrier constraining the electrons within the crystal.

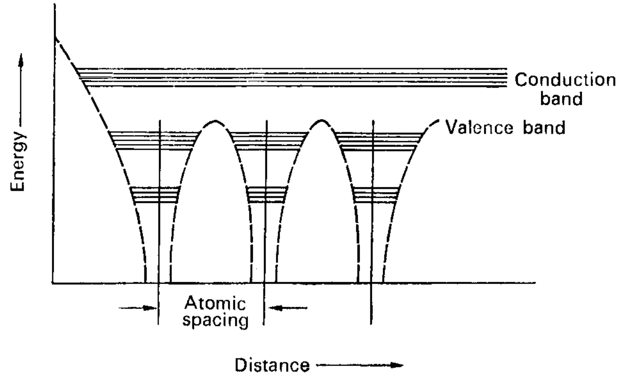


FIG. 2.1. Band structure.

The Bohr atom theory outlined in this section has been superseded for many purposes by the model developed using wave mechanics. This approach involves the relation between matter and radiation and de Broglie extended the theory to show that an electron could be represented by a wave motion. The wavelength λ associated with this wave motion is given by

$$\lambda = \frac{h}{m_e v},$$

where h is Planck's constant, m_e is the mass of an electron, and v is the velocity of an electron.

The theory is covered in detail in many physics textbooks, and leads to Schrodinger's wave equation the solution of which gives the probability of finding an electron at a given point.

2.2. Conductors

Conductors are characterized by having a number of free electrons able to wander through the conductor and be attracted to a positively applied potential. In terms of energy levels, this implies a partially filled conduction band as shown in Fig. 2.2.

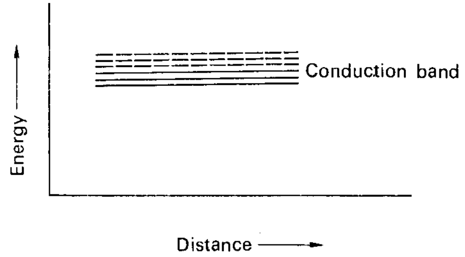


FIG. 2.2. Energy levels in a conductor.

Conduction takes place by electrons acquiring energy from the applied field and rising to a higher energy level. They drop back to lower energy levels on collision but immediately acquire more energy. There is a net movement or drift of electrons towards the positive electrode.

This drift velocity v_E is directly proportional to the applied field E. The constant of proportionality is known as the mobility of the electron μ_E;

i.e. $$v_E = \mu_E E.$$

If v_E is in m/s, E in V/m, the units of mobility are m²/Vs. The current I flowing at any instant may be equated to the charge passing a given point, per second;

i.e. $$I = N_E e v_E A,$$

where e is the charge on an electron, N_E is the number of free electrons/m^3, and A is the area of the conductor (m^2),

or
$$I = N_E e\mu_E EA$$
$$= N_E e\mu_E \frac{V}{L} A,$$

where V is the applied voltage and L the length of the conductor. Rearranging,
$$\frac{V}{I} = \frac{I}{N_E e\mu_E} \times \frac{L}{A}.$$

$1/(N_E e\mu_E)$ obviously corresponds to the resistivity ϱ. This increases with temperature due to a reduction in mobility resulting from the increased number of collisions caused by the larger atomic vibrations.

2.2.1 *Hall effect in conductors*

If a conductor carries a current I in a field of flux density B webers per square metre, the electrons will experience a deflecting force Bev_E newtons, as shown in Fig. 2.3.

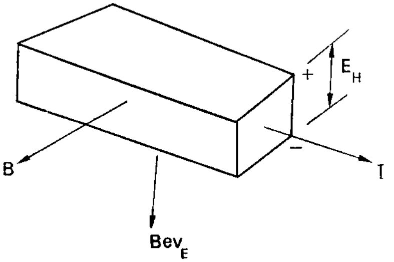


FIG. 2.3. Hall effect.

The accumulation of electrons at the base will give rise to an electric field E_H across the conductor and the production of this field is known as the Hall effect.

In equilibrium $$eE_H = Bev_E$$

or $$v_E = \frac{E_H}{B}.$$

The current density, $$J = \frac{I}{A} = N_E e v_E$$

$$= \frac{N_E e E_H}{B}.$$

Hence $1/N_E e = E_H/BJ$, and this is known as the Hall coefficient (R_H). This equation enables N_E to be determined.

Also

$$\mu_E = \frac{1}{N_E e \varrho} = \frac{R_H}{\varrho},$$

enabling the mobility to be found.

Inside the conductor the resultant electric field is due to two components—the applied field and the Hall field. The angle between these components is known as the Hall angle;

i.e. $$\tan\theta = \frac{E_H}{E}$$

$$= \frac{Bv_E}{E}$$

$$= B\mu_E \quad \text{since} \quad v_E = \mu_E E.$$

For practical purposes the Hall effect in conductors is very small. This follows from the equation $E_H = BJ/N_E e$, since the number of free electrons N_E for a conductor is very high, say 10^{29} per m^3.

2.3. Insulators

When the energy band diagram for an insulator is considered, there is a large energy gap between a full valence band and an empty conduction band. For conduction to take place electrons must acquire sufficient energy to jump this energy gap, and this leads to a decrease in resistivity with temperature. In practice, any conduction in insulators is usually due to minute traces of impurities.

2.4. Intrinsic Semiconductors

The energy band diagram for an intrinsic semiconductor is similar to that of an insulator but with a lower energy gap, 0·72 eV for germanium and 1·1 eV for silicon. This gap is such that even at room temperature some electrons are raised from the valence to the conduction band.

The total current flowing is then the sum of that due to electrons and holes;

i.e.
$$I = N_E e v_E A + N_H e v_H A,$$

where N_E and N_H are the number of electrons and holes respectively (obviously N_E and N_H are equal and may be put equal to N_I), v_E and v_H are the drift velocities of electrons and holes respectively, e is the charge on an electron, and A is the area of the semiconductor.

Also
$$v_E = \mu_E E \quad \text{and} \quad v_H = \mu_H E,$$

where μ_E and μ_H are the mobilities of electrons and holes respectively and E is the electric field strength.

Therefore
$$I = N_I e(\mu_E + \mu_H)\, AE,$$

but $E = V/L$, where V is the applied voltage and L the length of the semiconductor;

i.e.
$$I = N_I e(\mu_E + \mu_H)\frac{AV}{L}.$$

Rearranging,
$$\frac{V}{I} = \frac{I}{N_I e(\mu_E + \mu_H)} \times \frac{L}{A}.$$

Hence the resistivity of a sample of an intrinsic semiconductor is given by

$$\varrho = \frac{1}{N_I e(\mu_E + \mu_H)} \ \Omega\text{-m}.$$

To evaluate this expression, $N_I (= N_E = N_H)$ must be determined. This is a function of temperature and may be evaluated as in the following subsection.

2.4.1. Number of electrons in the conduction band

As previously stated, at any finite temperature some electrons will acquire sufficient thermal energy to raise them from the valence band to the conduction band. The actual number depends on the number of permissible electron energy levels and the probability of these levels being occupied.

Assume N_C is the number of electron energy levels per unit energy range (in the conduction band) and $f(e)$ is the probability of these levels being occupied. Then if E_C and E_T are the lower and upper limits of the conduction band, the number of electrons N_E in the conduction band is given by

$$N_E = \int_{E_Q}^{E_T} N_C f(e)\, dE.$$

In practice the distribution of the energy levels in the conduction band is such that the majority of them are close to E_C, and the top limit may then be taken as infinity,

i.e. $$N_E = \int_{E_Q}^{\infty} N_C f(e)\, dE.$$

The evaluation of N_E hence requires a knowledge of the density of electron states N_C and the probability $f(e)$.

(a) Density of electron states

To calculate the number of states, recourse has to be made to the quantum theory. This is based on momentum rather than energy considerations, and the fundamental hypothesis is that the smallest unit of momentum in one dimension is h, where h is Planck's constant. Hence in three dimensions the number of states that may be distinguished is $1/h^3$ per unit volume.

At low temperatures the valence band would be completely full and the conduction band completely empty. At room temperature some electrons acquire sufficient energy to cross the energy gap and reach the conduction band. Then if E_C is the lower limit of the conduction band, the total energy E may be written

$$\begin{aligned} E &= E_C + \text{kinetic energy} \\ &= E_C + \tfrac{1}{2} m_e v^2, \end{aligned}$$

where m_e is the effective mass of an electron and v is its velocity;

i.e. $$E = E_C + \frac{M^2}{2m_e},$$

where $M = m_e v$ is the momentum;

i.e. $$M = \sqrt{[2m_e(E - E_C)]}.$$

It has already been implied that the direction of the momentum is irrelevant, and hence for a given value of E the locus of M may be represented by a sphere of radius $M = \sqrt{[2m_e(E - E_C)]}$ as in Fig. 2.4.

If E increases by dE, the momentum will increase by dM, where

$$\begin{aligned} dM &= \sqrt{(2m_e)} \times \frac{1}{2} (E - E_C)^{-1/2}\, dE \\ &= \sqrt{\left[\frac{m_e}{2(E - E_C)}\right]}\, dE. \end{aligned}$$

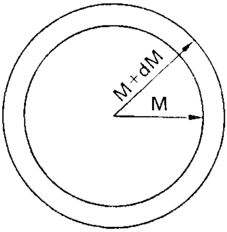


FIG. 2.4. Momentum shells.

The difference in volume between the two spheres is then $4\pi M^2\,dM$, so that for the conduction band the number of energy states in the momentum range dM (or energy range dE) is given by:

$$\text{number of energy states} = \frac{4\pi M^2\,dM}{h^3}$$

$$= \frac{4\pi}{h^3}\times 2m_e(E-E_C)\sqrt{\left[\frac{m_e}{2(E-E_C)}\right]}\,dE$$

$$= \frac{4\pi}{h^3}\sqrt{(2)}\,(m_e)^{3/2}\sqrt{(E-E_C)}\,dE.$$

Since two electrons can occupy each energy level (i.e. opposite spin), the number of electron states dNc in the energy range dE is given by

$$dNc = \frac{8\pi\sqrt{(2)}\,(m_e)^{3/2}}{h^3}\sqrt{(E-E_C)}\,dE,$$

i.e. number of electron energy levels per unit energy range is

$$Nc = \frac{dNc}{dE} = \frac{8\pi\sqrt{(2)}\,(m_e)^{3/2}}{h^3}\sqrt{(E-E_C)}.$$

The variation of energy states is then as shown in Fig. 2.5.

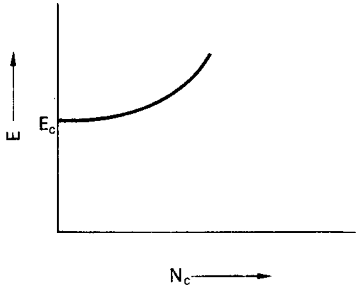


FIG. 2.5. Electron energy states.

(b) Probability of electron occupation

At low temperatures all energy levels are filled up to a certain value E_F. At any finite temperature some electrons will have higher energies and some of the lower energy levels will be empty. The probability of a certain energy level being filled follows the Fermi–Dirac distribution law,

i.e. $$f(e) = \frac{1}{1+\varepsilon^{[(E-E_F)/kT]}}.$$

From this equation it follows that at a certain value of $E = E_F$ (the Fermi level), the probability is always $\frac{1}{2}$ regardless of temperature, while at low energies $f(e)$ is unity, i.e. filled levels and at high energies $f(e)$ is zero, i.e. empty levels. Except at the Fermi level itself, temperature also has an effect, and the variation in probability is as shown in Fig. 2.6.

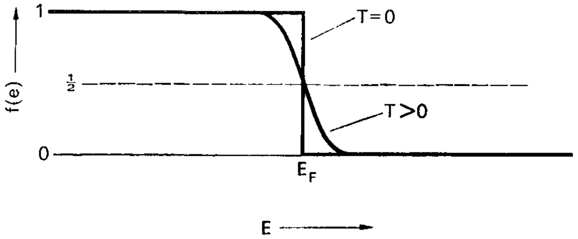


FIG. 2.6. Fermi–Dirac distribution.

For the conduction band, $E-E_F$ is considerably greater than kT, i.e. $\varepsilon^{[(E-E_F)/kT]}$ is much greater than unity and to a close approximation

$$f(e) = \varepsilon^{-[(E-E_F)/kT]}.$$

The number of electrons in the conduction band may now be estimated,

i.e.

$$\begin{aligned} N_E &= \int_{E_C}^{\infty} \frac{8\pi\sqrt{(2)}\,(m_e)^{3/2}}{h^3} \sqrt{(E-E_C)}\,\varepsilon^{-[(E-E_F)/kT]}\,dE \\ &= \frac{8\pi\sqrt{(2)}\,(m_e)^{3/2}}{h^3}\,\varepsilon^{-[(E_C-E_F)/kT]} \int_{E_C}^{\infty} \sqrt{(E-E_C)}\,\varepsilon^{-[(E-E_C)/kT]}\,dE. \end{aligned}$$

To solve this integral the substitution, $x = \sqrt{[(E-E_C)/kT]}$ may be used. Then the limits become 0 and ∞, and $dE = 2kTx\,dx$,

i.e. $$N_E = \frac{16\pi(kT)^{3/2}\sqrt{(2)}\,(m_e)^{3/2}}{h^3}\,\varepsilon^{-[(E_C-E_F)/kT]} \int_0^{\infty} x^2\varepsilon^{-x^2}\,dx.$$

Integrating by parts gives $\displaystyle\int_0^{\infty} x^2\varepsilon^{-x^2}\,dx = \frac{\sqrt{\pi}}{4}.$

Hence $$\begin{aligned} N_E &= \frac{2}{h^3}(2\pi m_e kT)^{3/2}\,\varepsilon^{-[(E_C-E_F)/kT]} \\ &= A\varepsilon^{-[(E_C-E_F)/kT]}, \end{aligned}$$

where $$A = \frac{2}{h^3}(2\pi m_e kT)^{3/2}.$$

2.4.2. Number of holes in the valence band

This derivation follows the same pattern as before. If E_V is the energy at the top of the valence band, the total energy E at a level within the valence band is given by

$$E = E_V - \frac{M^2}{2m_h},$$

where M is the momentum and m_h is the effective mass of a hole,

i.e. $$M = \sqrt{[2m_h(E_V - E)]}.$$

Hence the number of hole energy levels per unit energy range is given by

$$N_V = \frac{8\pi\sqrt{(2)}\,(m_h)^{3/2}}{h^3}\sqrt{(E_V - E)}.$$

The probability of an electron energy level being empty, i.e. the probability of a hole, is given by

$$f(h) = 1 - \frac{1}{1 + \varepsilon^{[(E-E_F)/kT]}} = \frac{\varepsilon^{[(E-E_F)/kT]}}{1 + \varepsilon^{[(E-E_F)/kT]}}.$$

But for the valence band, $(E_F - E)/kT$ is much greater than 1, i.e. $\varepsilon^{[(E-E_F)/kT]}$ is small. Hence the probability of a hole is given by

$$f(h) = \varepsilon^{-[(E_F-E)/kT]}.$$

The number of holes in the valence band is then

$$N_H = \int_{E_B}^{E_V} \frac{8\pi\sqrt{(2)}\,(m_h)^{3/2}}{h^3}\sqrt{(E_V - E)}\,\varepsilon^{-[(E_F-E)/kT]}\,dE,$$

where E_B corresponds to the bottom of the valence band. In practice this may be taken as zero.

The integration is then very similar to that carried out in determining the number of electrons in the conduction band and yields the result

$$N_H = \frac{2}{h^3}(2\pi m_h kT)^{3/2}\,\varepsilon^{-[(E_F-E_V)/kT]}$$

$$= B\varepsilon^{-[(E_F-E_V)/kT]},$$

where $$B = \frac{2}{h^3}(2\pi m_h kT)^{3/2}.$$

2.4.3. Fermi level for an intrinsic semiconductor

Obviously, for an intrinsic semiconductor the number of electrons in the conduction band must be equal to the number of holes in the valence band,

$$\text{i.e.}\quad \frac{2}{h^3}(2\pi m_e kT)^{3/2}\,\varepsilon^{-[(E_C-E_F)/kT]} = \frac{2}{h^3}(2\pi m_h kT)^{3/2}\,\varepsilon^{-[(E_F-E_V)/kT]}.$$

To a first approximation, $m_e = m_h$, and then

$$E_C-E_F = E_F-E_V \quad \text{or} \quad E_F = \tfrac{1}{2}(E_C+E_V).$$

This result shows that the Fermi level lies midway between the valence and conduction bands. The result may now be shown graphically as in Fig. 2.7.

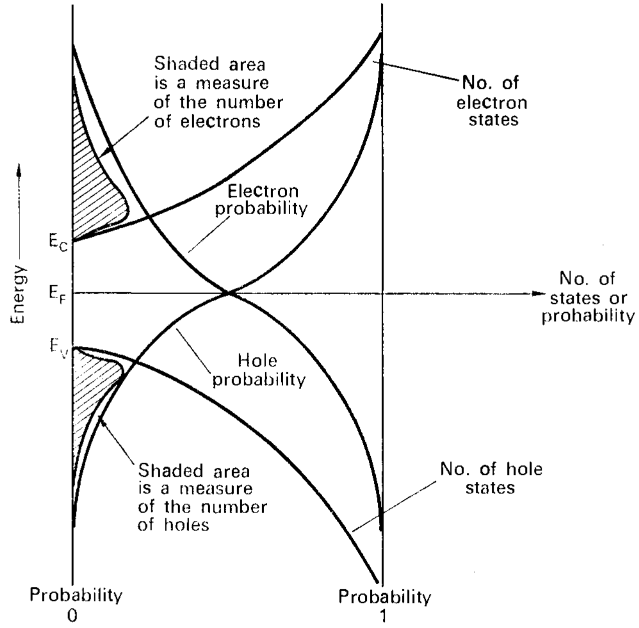

Shaded area is a measure of the number of electrons
No. of electron states
Electron probability
Energy
E_C
E_F
E_V
No. of states or probability
Hole probability
Shaded area is a measure of the number of holes
No. of hole states
Probability 0
Probability 1

FIG. 2.7. Electrons and holes in an intrinsic semiconductor.

If the value of E_F is now substituted, the number of electrons (or holes) becomes

$$\begin{aligned} N_E &= \frac{2}{h^3}(2\pi m_e kT)^{3/2}\,\varepsilon^{-\{[E_G-1/2(E_G+E_V)]/kT\}} \\ &= \frac{2}{h^3}(2\pi m_e kT)^{3/2}\,\varepsilon^{-[(E_G-E_V)/2kT]} \\ &= A\varepsilon^{-[(E_G-E_V)/2kT]}, \end{aligned}$$

where

$$A = \frac{2}{h^3}(2\pi m_e kT)^{3/2}.$$

This equation leads to the following results:

(1) The smaller the energy gap the larger the number of "free" holes and electrons, i.e. intrinsic germanium has a larger number of carriers than intrinsic silicon at a given temperature.

(2) The higher the temperature the larger the number of holes and electrons.

Also

$$N_E N_H = N_I^2 = \frac{4}{h^6}(2\pi kT)^3 (m_e m_h)^{3/2}\,\varepsilon^{-[(E_G-E_V)/kT]}.$$

2.4.4. *Diffusion currents in an intrinsic semiconductor*

In section 2.4.1 the current that was considered was the drift of carriers due to an applied field. In general this is only one of two components of the current flowing in a semiconductor, the other is that due to diffusion. The latter occurs without the application of an applied field and is the natural tendency of concentrations of carriers to spread out uniformly through a sample of semiconductor material.

If light is focused on one spot in a semiconductor, hole–electron pairs will be produced, and these will tend to spread out uniformly throughout the material. The corresponding movement of charge constitutes a current, and the magnitude of this current is proportional to the concentration gradient, i.e. considering electrons flowing in one direction,

$$I_E = \text{constant} \times \text{electron charge} \times \text{area} \times \text{concentration gradient}$$

$$= \text{constant} \times eA \frac{dN_E}{dx},$$

where A is the area, e is the charge on an electron, and dN_E/dx is the electron concentration gradient.

The constant is known as the diffusion coefficient, denoted by D_E,

i.e. $$I_E = D_E eA \frac{dN_E}{dx}.$$

Similarly for holes,

$$I_H = D_H eA \frac{dN_H}{dx}.$$

In practice, since electrons and holes carry opposite charges, the sign of one of these expressions should be negative. Taking conventional current as the positive direction, the minus sign would appear in the expression for the hole flow.

The results obtained previously allow expressions for the total current flowing in a sample of intrinsic semiconductor material to be stated when both drift and diffusion currents are included. If the sample of intrinsic material shown in Fig. 2.8 is considered, the components of drift and diffusion current will be as indicated. The total electron current is given by

$$I_E = N_E e \mu_E AE + D_E eA \frac{dN_E}{dx},$$

and the total hole current is represented by

$$I_H = N_H e \mu_H AE - D_H eA \frac{dN_H}{dx}.$$

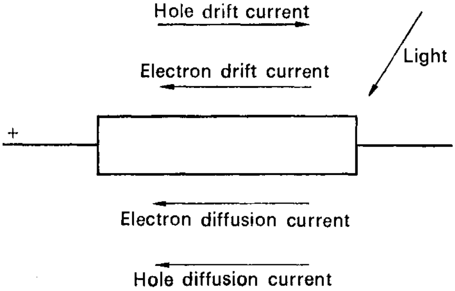


FIG. 2.8. Drift and diffusion currents.

Since the number of holes and electrons are the same, $N_E = N_H = N_I$. Hence the total current I is given by

$$I = N_I e(\mu_E + \mu_H)AE + \left(D_E \frac{dN_E}{dx} - D_H \frac{dN_H}{dx}\right)eA.$$

In fact, μ_E, μ_H, D_E and D_H are directly related by the Einstein relation,

i.e. $$\frac{D_H}{\mu_H} = \frac{D_E}{\mu_E} = \frac{kT}{e}.$$

Some connection between mobility and diffusion constant is to be expected since they are both controlled by collisions.

2.5. Extrinsic Semiconductors

The introduction of pentavalent or trivalent impurities in intrinsic germanium or silicon in very small amounts, normally about 1 part in 10^8, has a large effect on the conductivity.

The energy level diagram for extrinsic semiconductor with added impurities are shown in Figs. 2.9 and 2.10. Figure 2.9 shows the energy band diagram for a *n*-type semiconductor with an added pentavalent impurity such as phosphorus, arsenic, or antimony. Figure 2.10 shows the corresponding diagram for a *p*-type semi-

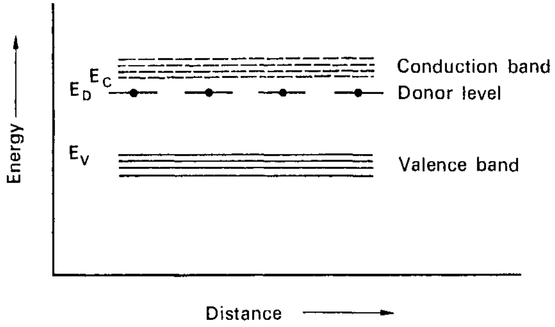


FIG. 2.9. Energy levels for an *n*-type semiconductor.

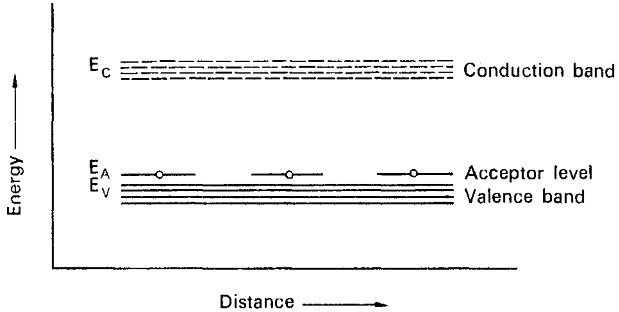


FIG. 2.10. Energy levels for a *p*-type semiconductor.

conductor with an added trivalent impurity such as boron, aluminium, gallium, or indium.

If an isolated piece of extrinsic semiconductor material, where either the electrons or holes predominate, is considered, there is in general some random movement of charge carriers. This gives rise to drift and diffusion currents, but since there is no net movement of carriers both the total electron current and the total hole current must be zero;

i.e. $$I_E = N_E \mu_E eAE + D_E eA\frac{dN_E}{dx} = 0$$

and $$I_H = N_H \mu_H eAE - D_H eA\frac{dN_H}{dx} = 0.$$

From these equations,

$$N_E \mu_E E = -D_E \frac{dN_E}{dx} \quad \text{and} \quad N_H \mu_H E = D_H \frac{dN_H}{dx}.$$

Dividing and rearranging, $$-\frac{dN_E/dx}{N_E} \times \frac{D_E}{\mu_E} = \frac{dN_H/dx}{N_H} \times \frac{D_H}{\mu_H}.$$

But from the Einstein relation,

$$\frac{D_E}{\mu_E} = \frac{D_H}{\mu_H}.$$

Hence $$-\frac{dN_E}{N_E} = \frac{dN_H}{N_H}.$$

Integrating, $$-\log_e N_E = \log_e N_H + K$$

or $$N_E N_H = K_1,$$

where K_1 is a constant.

No limitations have been placed on the values of N_E and N_H, and the expression is valid for intrinsic material. Hence the constant K_1 must be equal to N_I^2. This shows that for any extrinsic material the number of electrons N_E and holes N_H is given by

$$N_E N_H = N_I^2,$$

where N_I is the number of holes or electrons before doping.

This equation is an algebraic statement of the law of mass action. Physically, the introduction of an *n*-type impurity into an intrinsic material fills some of the holes produced by thermal action, while the addition of a *p*-type impurity into an intrinsic material absorbs some of the thermally produced electrons.

The law of mass action enables the resistivity of extrinsic semiconductors to be estimated. As in the case of intrinsic material, diffusion currents are not considered when the resistivity is evaluated since there is a negligible variation of charge concentration along the sample.

Consider the case of a sample of intrinsic germanium with 2×10^{19} charge carriers per m^3 at a given temperature. If 10^{20} atoms per m^3 of indium, a trivalent impurity, are added,

$$\begin{aligned} N_E N_H &= N_I^2 \\ &= 4\times 10^{38}. \end{aligned}$$

But since 10^{20} atoms of impurity N_A have been added, the difference between the number of holes and electrons is 10^{20} per m^3;

i.e. $$N_H - N_E = 10^{20}.$$

Solving these expressions,

$$N_H - \frac{4\times 10^{38}}{N_H} = 10^{20}$$

or $$N_H^2 - 10^{20}\, N_H - 4\times 10^{38} = 0.$$

Solving and taking the positive root,

$$N_E = 1{\cdot}04\times 10^{20}.$$

Hence, $$N_H = 0{\cdot}04\times 10^{20}.$$

The current flowing through the sample is then given by

$$\begin{aligned} I &= \text{current due to electrons} + \text{current due to holes} \\ &= N_E e \mu_E AE + N_H e \mu_H AE \\ &= e(N_E \mu_E + N_H \mu_H)AE, \end{aligned}$$

where A is the area of the sample, E is the applied field, e is the charge on an electron, N_E and N_H are the number of "free" electrons and holes, and μ_E and μ_H are the mobilities of electrons and holes, i.e. 0·36 m^2/V sec and 0·17 m^2/V sec respectively.

Substituting values

$$I = 1{\cdot}6\times 10^{-19}\,(0{\cdot}04\times 10^{20}\times 0{\cdot}36 + 1{\cdot}04\times 10^{20}\times 0{\cdot}17)AE$$

taking unit dimensions $A = 1$, $E = V$.

Therefore resistivity

$$\frac{V}{I} = \frac{1}{1{\cdot}6\times10^{-19}(0{\cdot}04\times10^{20}\times0{\cdot}36+1{\cdot}04\times10^{20}\times0{\cdot}17)}$$

$$= \underline{0{\cdot}327\ \Omega\text{-m.}}$$

The addition of a further 2×10^{20} atoms/m^3 of antimony (N_D), a pentavalent impurity, will swamp the effects of the trivalent impurity.

Then $$N_EN_H = N_I^2 = 4\times10^{38},$$

$$N_E-N_H = N_D-N_A = 2\times10^{20}-10^{20}$$

$$= 10^{20}.$$

Solving as before gives

$$N_E = 1{\cdot}04\times10^{20} \quad \text{and} \quad N_H = 0{\cdot}04\times10^{20},$$

and hence resistivity $\varrho = 0{\cdot}164\ \Omega$-m.

2.5.1. *Fermi level for an n-type semiconductor*

When an intrinsic semiconductor is doped with a pentavalent or donor impurity, one of the electrons of each impurity atom lies outside the crystal structure. It may be classed as a "free" electron, but due to the higher positive charge on the impurity atom there is a slight tendency for the electron to remain in the neighbourhood of that atom. Its energy level E_D is slightly below that of the conduction band, being about 0·01 eV as indicated in Fig. 2.9. The density of electron states must be the same as for an intrinsic material, but since, at any finite temperature, electrons can easily be raised from the donor level into the conduction band, the number of electrons available for conduction is normally much greater than the number of holes. The equation for the probability function is the same, but in order to satisfy the conditions stated, the value of the Fermi level E_{F_N} must rise. This is shown diagrammatically in Fig. 2.11.

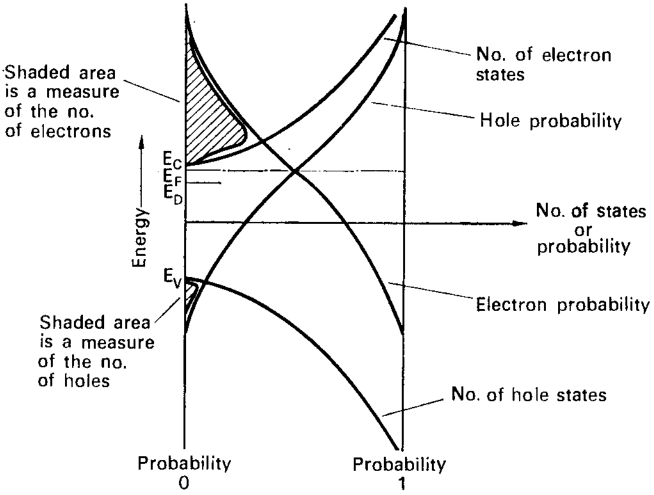


FIG. 2.11. Electrons and holes in an *n*-type semiconductor.

Then using the relation

$$N_E - N_H = N_D,$$

where N_E, N_H, and N_D are the numbers of electrons, holes, and donor atoms respectively,

$$A\varepsilon^{-[(E_C-E_{F_N})/kT]} - B\varepsilon^{-[(E_{F_N}-E_V)/2kT]} = N_D,$$

where $$A = \frac{2}{h^3}(2\pi m_e kT)^{3/2} = N_I\varepsilon^{[(E_C-E_V)/2kT]},$$

$$B = \frac{2}{h^3}(2\pi m_h kT)^{3/2} = N_I\varepsilon^{[(E_C-E_V)/2kT]}.$$

Therefore $$N_I\varepsilon^{[(2E_{F_N}-E_C-E_V)/2kT]} - N_I\varepsilon^{-[(2E_{F_N}-E_C-E_V)/2kT]} = N_D$$

or $$2N_I \sinh\left(\frac{2E_{F_N}-E_C-E_V}{kT}\right) = N_D,$$

i.e. $$E_{F_N} = \frac{1}{2}(E_C+E_V) + kT\sinh^{-1}\frac{N_D}{2N_I}.$$

This expression shows that the Fermi level is above that of an intrinsic semiconductor by an amount $kT\sinh^{-1}\left(\frac{N_D}{2N_I}\right)$. At high temperatures, where N_I is much greater than N_D, the Fermi level reverts to its intrinsic value of $\frac{1}{2}(E_C+E_V)$.

The above derivation does, in fact, assume that all the impurity atoms are ionized. To allow for unionized donors, the probability for the donors to be ionized should be included. This involves estimating the probability of finding a hole at any energy level E_D and means that N_D should be replaced by

$$\frac{N_D}{1+2\varepsilon^{-[(E_D-E_{F_N})/kT]}}.$$

2.5.2. *Fermi level for a p-type semiconductor*

The introduction of a trivalent or acceptor impurity gives rise to empty energy levels E_A just above the valence band as indicated in Fig. 2.10. The density of states must be the same as for an intrinsic material but since, at any finite temperature, electrons can easily be raised from the valence band to the acceptor level, the number of holes available for conduction is normally much greater than the number of electrons. The equation for the probability function is the same, but in order to satisfy the conditions stated the value of the Fermi level (E_{F_P}) must fall. This is shown diagrammatically in Fig. 2.12.

Then using the relation

$$N_H-N_E = N_A,$$

where N_H, N_E, and N_A are the numbers of holes, electrons, and acceptor atoms respectively,

$$B\varepsilon^{-[(E_{F_P}-E_V)/kT]}-A\varepsilon^{-[(E_C-E_{F_P})/kT]} = N_A,$$

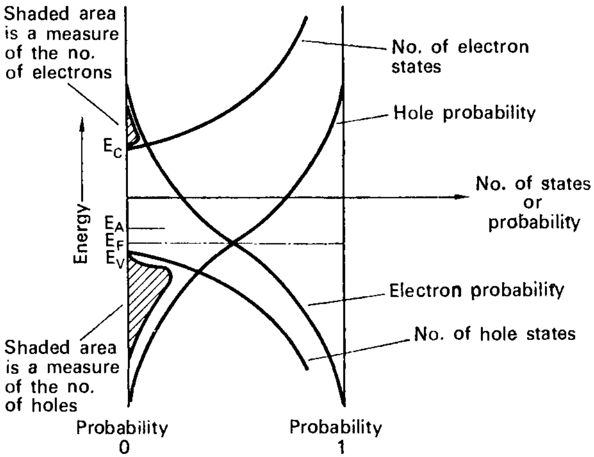


FIG. 2.12. Electrons and holes in a *p*-type semiconductor.

where
$$B = \frac{2}{h^3}(2\pi m_h kT)^{3/2} = N_I \varepsilon^{[(E_C-E_V)/2kT]},$$

$$A = \frac{2}{h^3}(2\pi m_e kT)^{3/2} = N_I \varepsilon^{[(E_C-E_V)/2kT]},$$

i.e.
$$N_I \varepsilon^{[(E_C+E_V-2E_{F_P})/2kT]} - N_I \varepsilon^{-[(E_C+E_V-2E_{F_P})/2kT]} = N_A$$

or
$$2N_I \sinh\left(\frac{E_C+E_V-2E_{F_P}}{2kT}\right) = N_A,$$

i.e.
$$E_{F_P} = \frac{1}{2}(E_C+E_V) - kT\sinh^{-1}\left(\frac{N_A}{2N_I}\right).$$

This expression shows that the Fermi level is below that of an intrinsic semiconductor by an amount $kT\sinh^{-1}\left(\frac{N_A}{2N_I}\right)$. At high temperature, where N_I is much greater than N_A, the Fermi level reverts to its intrinsic value of $\frac{1}{2}(E_C+E_V)$.

As in the case of *n*-type material, this derivation assumes that all the holes in the acceptor level are filled and a more accurate result could be obtained in the probability of finding an electron at an energy level E_A is estimated.

2.5.3. Hall effect in extrinsic semiconductors

The Hall effect is more noticeable with extrinsic semiconductors than conductors, but in this case the Hall coefficient R_H is usually taken as $\frac{3\pi}{8}\times\frac{1}{N_E e}$ m³/C for an *n*-type semiconductor, and $-\frac{3\pi}{8}\times\frac{1}{N_H e}$ m³/C for a *p*-type semiconductor. In these expressions, N_E and N_H are the number of "free" electrons and "free" holes respectively. The factor $\frac{3\pi}{8}$ is due to a different distribution of velocities of the carriers in a semiconductor and must be modified for large values of field strength.

The values of Hall e.m.f. for semiconductors may be relatively high since the number of "free" electrons or holes is relatively low, say 10^{21} per m³ compared with 10^{29} per m³ for a conductor.

Questions on Chapter 2

Q.1. Calculate the radii and potential energy levels of the first two possible orbits for an ionized helium atom.

Q.2. The number of free electrons in copper is $1{\cdot}1\times10^{29}$ per m³ and its resistivity is 10^{-8} Ω-m. Determine the mobility of the electrons and the Hall coefficient for copper. Hence determine the current density which gives a Hall voltage of 1 μV in a field of 1 tesla.

Q.3. Explain what is meant by the term mobility and how it is related to the total current flowing in a piece of intrinsic semiconductor material.

The resistivity of pure germanium at a particular temperature is 0·52 Ω-m. If it is doped with 10^{20} atoms/m³ of a trivalent impurity, estimate the new value of resistivity. Assume the mobility of holes and electrons are 0·2 and 0·4 m²/V s respectively, and the charge on an electron is $1{\cdot}6\times10^{-19}$ C.

Q.4. Explain what is meant by the terms Fermi level, density of states, and Fermi–Dirac distribution law. Illustrate your answer by sketches showing how the electron and hole densities vary in intrinsic, *p*-type, and *n*-type semiconductors.

Q.5. Show that in the case of a conductor the Hall coefficient is given by

$$R_H = \frac{1}{N_E e} = \frac{E_H}{BJ}.$$

How is this expression modified in the case of a semiconductor? A sample of *n*-type germanium is 1 cm square and 0·5 cm thick. When a current of 5 mA flows from a 1·35 V supply through a narrow face and there is a transverse field of 0·045 tesla, the Hall voltage developed is 20 mV. Determine the Hall coefficient and electron mobility for germanium.

CHAPTER 3

Physics of Semiconductor Devices

IN THIS chapter the physical concepts introduced in the previous chapter will be extended to cover complete devices. The *pn* junction is the basis of all the devices and will initially be considered in detail. Later, the results obtained will be applied to multi-junction devices.

3.1. The *pn* Junction

The rectifying action of a *pn* junction depends on diffusion currents which flow without the application of an external field. In the formation of a junction, diffusion of electrons from the *n* type and holes from the *p* type give rise to a central depletion region produced by the recombination of holes and electrons. The diffusion of carriers continues until the charge transfer is sufficient to build up an energy barrier high enough to stop further migration of carriers. Any minority carriers (electrons in the *p* type and holes in the *n* type) produced thermally are swept across the junction as a drift current but are immediately balanced by a diffusion current due to the flow of majority carriers.

In terms of the energy levels diagram, electrons flow into the *p*-type region until the raising of its energy level diagram, coupled with a fall in that of the *n*-type region, produces a common Fermi level as shown in Fig. 3.1.

The energy barrier must be equal to the difference of the Fermi levels of the *p* and *n* regions if they were considered in isolation.

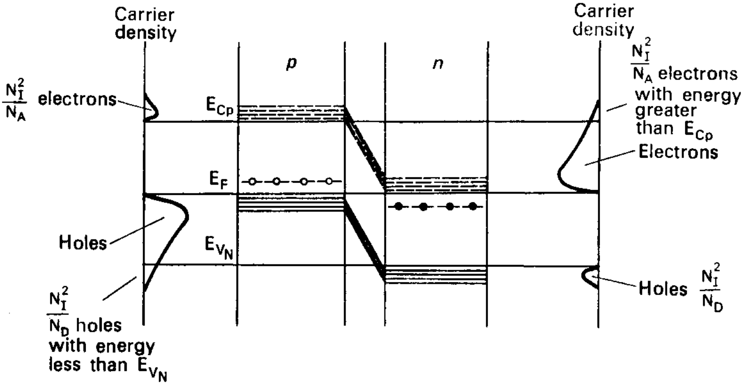

Carrier density
$\frac{N_I^2}{N_A}$ electrons
Holes
$\frac{N_I^2}{N_D}$ holes with energy less than E_{V_N}
E_{C_P}
E_F
E_{V_N}
p
n
Carrier density
$\frac{N_I^2}{N_A}$ electrons with energy greater than E_{C_P}
Electrons
Holes $\frac{N_I^2}{N_D}$

FIG. 3.1. *pn* junction with zero bias.

If V_B is voltage developed across the junction,

$$\begin{aligned}\text{energy barrier} &= eV_B\\ &= E_{F_N}-E_{F_P}\\ &= \left[\frac{1}{2}(E_C+E_V)+kT\sinh^{-1}\frac{N_D}{2N_I}\right]\\ &\quad -\left[\frac{1}{2}(E_C+E_V)-kT\sinh^{-1}\frac{N_A}{2N_I}\right]\\ &= kT\left(\sinh^{-1}\frac{N_D}{2N_I}+\sinh^{-1}\frac{N_A}{2N_I}\right).\end{aligned}$$

Under normal conditions, N_D and N_A are much greater than N_I and to a close approximation

$$\sinh^{-1}\frac{N_D}{2N_I}=\log_\varepsilon\left\{\frac{N_D}{2N_I}+\sqrt{\left[\left(\frac{N_D}{2N_I}\right)^2+1\right]}\right\}\simeq\log_\varepsilon\frac{N_D}{N_I}$$

and $$\sinh^{-1}\frac{N_A}{2N_I}\simeq\log_\varepsilon\frac{N_A}{N_I}.$$

Therefore
$$eV_B = kT\left(\log_\varepsilon \frac{N_D}{N_I} + \log_\varepsilon \frac{N_A}{N_I}\right)$$
$$= kT \log_\varepsilon \frac{N_D N_A}{N_I^2}. \qquad (3.1.1)$$

This expression gives the height of the energy barrier with no applied bias.

Under these equilibrium conditions, the number of electrons with energies greater than E_{CP} (the bottom of the conduction band of the *p*-type region) are equal for both *p*- and *n*-type regions. In a similar manner, the number of holes with energies less than E_{VN} (the top of the valence band of the *n*-type region) is the same for both *p* and *n* regions. Such conditions are illustrated in Fig. 3.1.

The number of thermally generated electron carriers in the *p*-type region is given by $\frac{N_I^2}{N_A}$ [$= N_D \varepsilon^{-eV_B/kT}$ from eqn. (3.1.1)] and must be equal to the number of electrons that contribute to a diffusion current from the *n*- to the *p*-type regions. In practice, once these carriers enter the *p* region their numbers fall exponentially as shown in Fig. 3.2. This is due to recombination, but the initial rate of fall is $N_I^2/(N_A L_N)$, where L_N is the electron diffusion length, defined as the average distance an electron travels before recombination.

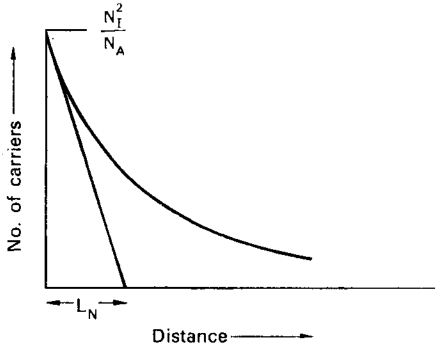


FIG. 3.2. Diffusion length for minority carriers.

Hence under equilibrium conditions,
electron drift current from *p* to *n*

$$= \text{electron diffusion current from } n \text{ to } p$$

$$= \frac{D_E e A N_I^2}{N_A L_N}.$$

The movement of holes under equilibrium conditions may be considered in the same fashion;
hole drift current from *n* to *p*

$$= \text{hole diffusion current from } p \text{ to } n$$

$$= \frac{D_H e A N_I^2}{N_D L_P},$$

where L_P is the hole diffusion length.

3.1.1. *Current flow with reverse bias*

When the *p* side of the junction is made negative with respect to the *n* side, the energy barrier is increased by an amount eV_R joules, where V_R is the applied voltage. The energy band diagram is shown in Fig. 3.3, the width of the depletion layer itself increasing

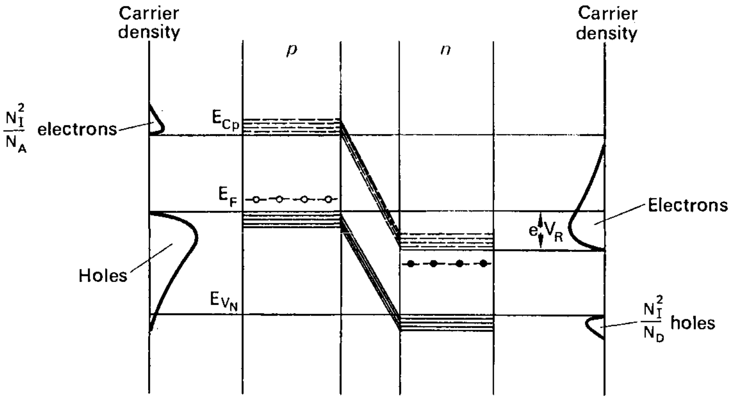


FIG. 3.3. *pn* junction with reverse bias.

slightly. In this case the number of carriers that can contribute to a diffusion current across the junction is the very small number of electrons in the n region with energies greater than E_{CP} and the very small number of holes in the p region with energies less than E_{VN}. There is, however, a drift current still flowing across the junction due to electrons produced thermally in the p region and holes produced thermally in the n region.

The magnitude of this current has already been evaluated for equilibrium conditions and is known as the reverse saturation current I_S. It is independent of the magnitude of the reverse voltage and given by

$$\begin{aligned} I_S &= \frac{D_E e A N_I^2}{N_A L_N} + \frac{D_H e A N_I^2}{N_D L_P} \\ &= e A N_I^2 \left(\frac{D_E}{N_A L_N} + \frac{D_H}{N_D L_P} \right). \end{aligned}$$

The reverse current is reduced slightly for very low values of reverse voltage when a small diffusion current flows.

3.1.2. *Current flow with forward bias*

When the p side of the junction is made positive with respect to the n side, the energy barrier is lowered by an amount eV_F joules, where e is the charge on an electron and V_F the voltage applied. The energy level diagram for this condition is shown in Fig. 3.4, the width of the depletion region itself being reduced.

In this case the number of electrons that can contribute to a diffusion current is proportional to the number of electron carriers in the n-type regions with energy greater than E_{CP}. This number is relatively large, as is the number of holes with energy less than E_{VN} that can contribute to a hole diffusion current. Hence the total diffusion current for a forward-biased junction is relatively large, while the reverse drift current remains at the same value as that for an unbiased junction.

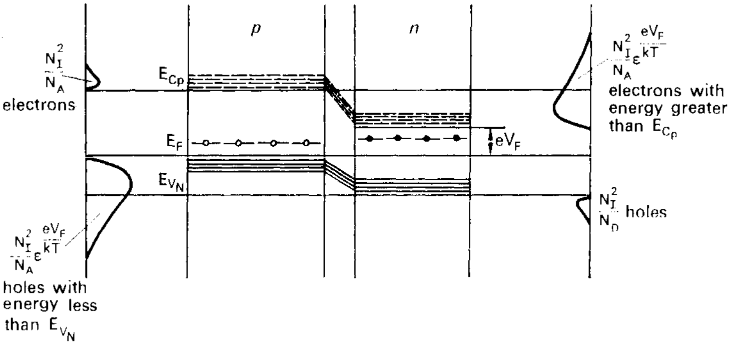

$\frac{N_I^2}{N_A}$ electrons
p
n
E_{C_P}
E_F
E_{V_N}
eV_F
$\frac{N_I^2}{N_A}\varepsilon^{\frac{eV_F}{kT}}$ electrons with energy greater than E_{C_P}
$\frac{N_I^2}{N_D}$ holes
$\frac{N_I^2}{N_A}\varepsilon^{\frac{eV_F}{kT}}$ holes with energy less than E_{V_N}

FIG. 3.4. *pn* junction with forward bias.

It has already been shown that for equilibrium conditions the number of electrons contributing to the diffusion current at the boundary between the depletion region and the *p* type is $\frac{N_I^2}{N_A} = N_D\varepsilon^{-[(eV_B)/kT]}$. With a forward bias, non-equilibrium conditions result giving rise to an increased number of electrons contributing to a diffusion current. At the boundary between the depletion region and the *p* type the number of such electrons is $N_D\varepsilon^{\{[e(V_B-V_F)]/kT\}} = \frac{N_I^2}{N_A}\varepsilon^{(eV_F)/kT}$. The initial rate of fall is then $\frac{N_I^2}{N_AL_N}\varepsilon^{(eV_F)/kT}$ and gives rise to an electron diffusion current of $\frac{DeAN_I^2}{N_AL_N}\varepsilon^{(eV_F)/kT}$.

In a similar manner the hole diffusion current at the boundary between the depletion region and the *n* type is given by $\frac{D_HeAN_I^2}{N_DL_P}\varepsilon^{(eV_F)/kT}$.

Since the amount of recombination in the depletion region is negligible, the total diffusion current is then

$$eAN_I^2\left(\frac{D_E}{N_AL_N}+\frac{D_H}{N_DL_P}\right)\varepsilon^{(eV_F)/kT}.$$

Subtracting the reverse drift current $eAN_I^2\left(\frac{D_E}{N_A L_N}+\frac{D_H}{N_D L_P}\right)$ enables the forward current I_F to be stated,

$$\text{i.e.} \qquad I_F = eAN_I^2\left(\frac{D_E}{N_A L_N}+\frac{D_H}{N_D L_P}\right)(\varepsilon^{(eV_F)/kT}-1)$$

$$= I_S(\varepsilon^{(eV_F)/kT}-1),$$

where I_S is the reverse saturation current.

If the suffix is dropped,

$$I = I_S(\varepsilon^{eV/kT}-1).$$

This expression then represents the complete characteristic shown in Fig. 3.5, since for negative values of V, $\varepsilon^{eV/kT}$ is negligible and $I = -I_S$.

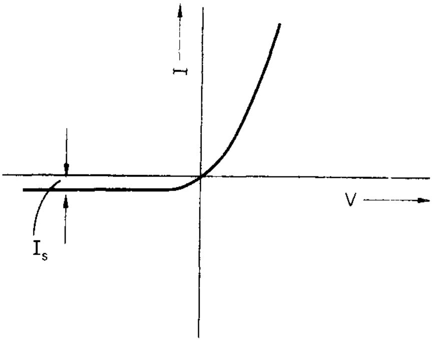


FIG. 3.5. Characteristic of a *pn* junction.

3.2. The Transistor

The simplest representation of a transistor is two diodes back to back, the first, which is forward biased, providing minority carriers for the second reverse-biased diode. It is a logical step to derive relations representing the operation of the transistor based on the diode equation $I = I_S(\varepsilon^{(eV/kT)}-1)$. Such equations were first derived by J. J. Ebers and J. L. Moll in the now historic paper published in the *Proceedings of the IRE* (December 1954).

Since both diodes provide minority carriers and each affects the current flowing through the other, the emitter and collector currents for a *pnp* transistor may be written in the form

$$I_E = I_1(\varepsilon^{(eV_{EB})/kT}-1)+I_2(\varepsilon^{(eV_{CB})/kT}-1) \tag{3.2.1}$$

and

$$I_C = I_3(\varepsilon^{(eV_{EB})/kT}-1)+I_4(\varepsilon^{(eV_{CB})/kT}-1). \tag{3.2.2}$$

For a *npn* device the voltage subscripts must be altered. In these expressions, I_1, I_2, I_3, and I_4 correspond to the reverse saturation current of a diode but are not those of the diodes considered separately. The common-base connection is assumed and V_{EB} and V_{CB} are the emitter–base and collector–base voltages respectively.

The main assumptions made in these relations are that the only voltage drops within the device are at the junctions themselves, i.e. the resistivity of the semiconductor material is low and the injected current densities are small.

The currents I_1, I_2, I_3, and I_4 may be evaluated in terms of the normal and inverted (i.e. with the roles of emitter and collector reversed) current gains and the corresponding leakage currents. These are:

h_{FB}	the d.c. or large-signal forward current gain with the emitter and collector biased normally.
h_{FBI}	The d.c. or large-signal current gain with the roles of emitter and collector reversed.
I_{CO}	the collector (leakage) current flowing when the emitter is open circuited and the collector junction is reverse biased.
I_{EO}	the emitter (leakage) current flowing when the collector is open circuited and the emitter junction is reverse biased.

If the collector junction is considered under normal bias conditions with V_{CB} relatively large and negative, the term $\varepsilon^{(eV_{CB})/kT}$ is much less than unity,

i.e.
$$I_E = I_1(\varepsilon^{(eV_{EB})/kT}-1)-I_2, \qquad (3.2.3)$$

$$I_C = I_3(\varepsilon^{(eV_{EB})/kT}-1)-I_4. \qquad (3.2.4)$$

Therefore
$$\frac{I_E+I_2}{I_1} = \frac{I_C+I_4}{I_3}$$

or
$$I_C = \frac{I_3}{I_1}\times I_E+\left(\frac{I_2I_3}{I_1}-I_4\right).$$

This expression may be compared with that expected for the forward characteristic shown in Fig. 3.6,

i.e.
$$I_C = h_{FB}I_E+I_{CO},$$

where h_{FB} and I_{CO} are negative.

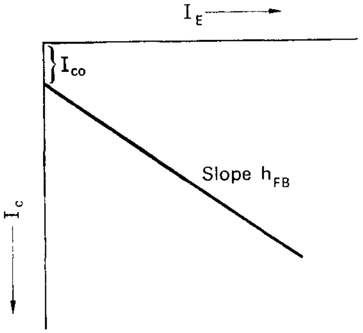


FIG. 3.6. Forward transfer characteristic of a *pnp* transistor.

Comparing coefficients,

$$h_{FB} = \frac{I_3}{I_1} \quad \text{and} \quad I_{CO} = I_4\left(\frac{I_2I_3}{I_1I_4}-1\right).$$

When the roles of emitter and collector are reversed, i.e. emitter reverse biased such that $\varepsilon^{(eV_{EB})/kT}$ is much less than unity,

$$I_E = I_2(\varepsilon^{(eV_{CB})/kT}-1)-I_1,$$

$$I_C = I_4(\varepsilon^{(eV_{CB})/kT}-1)-I_3,$$

giving
$$I_E = \frac{I_2}{I_4}\times I_C+\left(\frac{I_2I_3}{I_4}-I_1\right).$$

Hence comparing coefficients with the expression $I_E = h_{FBI}I_C+I_{EO}$ (Fig. 3.7),

$$h_{FBI} = \frac{I_2}{I_4} \quad \text{and } I_{EO} = I_1\left(\frac{I_2I_3}{I_1I_4}-1\right).$$

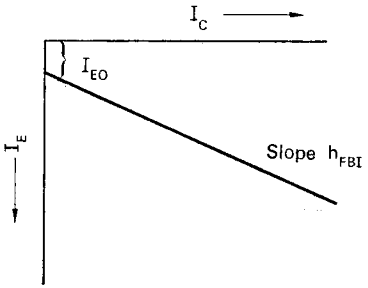


FIG. 3.7. Inverse transfer characteristic of a *pnp* transistor.

Therefore
$$\frac{I_2I_3}{I_1I_4} = h_{FB}h_{FBI},$$

giving
$$I_4 = -\frac{I_{CO}}{1-h_{FB}h_{FBI}}$$

and
$$I_1 = -\frac{I_{EO}}{1-h_{FB}h_{FBI}}.$$

Also
$$I_2 = -\frac{h_{FBI}I_{CO}}{1-h_{FB}h_{FBI}}$$

and
$$I_3 = -\frac{h_{FB}I_{EO}}{1-h_{FB}h_{FBI}}.$$

The currents I_2 and I_3 in fact represent a "coupling" term between the emitter and collector junctions. It is reasonable to assume these are equal. Therefore

$$h_{FBI}I_{CO} = h_{FB}I_{EO}. \tag{3.2.5}$$

The complete expressions may now be stated.

$$I_E = -\frac{I_{EO}}{(1-h_{FB}h_{FBI})}(\varepsilon^{(eV_{EB})/kT}-1) - \frac{h_{FBI}I_{CO}}{(1-h_{FB}h_{FBI})}(\varepsilon^{(eV_{CB})/kT}-1) \tag{3.2.6}$$

$$I_C = -\frac{h_{FB}I_{EO}}{(1-h_{FB}h_{FBI})}(\varepsilon^{(eV_{EB})/kT}-1) - \frac{I_{CO}}{(1-h_{FB}h_{FBI})}(\varepsilon^{(eV_{CB})/kT}-1). \tag{3.2.7}$$

Useful variations of these expressions may be obtained by eliminating one of the brackets containing an exponential term,

i.e. $$I_E-h_{FBI}I_C = -\frac{I_{EO}}{(1-h_{FB}h_{FBI})}(\varepsilon^{(eV_{EB})/kT}-1)(1-h_{FB}h_{FBI})$$

$$= -I_{EO}(\varepsilon^{(eV_{EB})/kT}-1) \tag{3.2.8}$$

and $$I_C-h_{FB}I_E = -I_{CO}(\varepsilon^{(eV_{CB})/kT}-1). \tag{3.2.9}$$

A d.c. equivalent circuit for the transistor under normal operating conditions, i.e. $\varepsilon^{(eV_{CB})/kT}$ negligible compared with unity, may now be derived from the relation $I_C = h_{FB}I_E + I_{CO}$ and eqn. (3.2.3):

$$I_E = I_1(\varepsilon^{(eV_{EB})/kT}-1)-I_2$$

$$= -\frac{I_{EO}}{(1-h_{FB}h_{FBI})}(\varepsilon^{(eV_{EB})/kT}-1)+\frac{h_{FBI}I_{CO}}{(1-h_{FB}h_{FBI})}$$

$$= -\frac{I_{EO}}{(1-h_{FB}h_{FBI})}(\varepsilon^{(eV_{EB})/kT}-1)+\frac{h_{FB}I_{EO}}{(1-h_{FB}h_{FBI})},$$

since $\qquad h_{FBI}I_{CO} = h_{FB}I_{EO} \quad$ from eqn. (3.2.5)

or
$$I_E = -\frac{I_{EO}}{(1-h_{FB}h_{FBI})}\varepsilon^{(eV_{EB})/kT} + \frac{(1+h_{FB})I_{EO}}{(1-h_{FB}h_{FBI})}. \qquad (3.2.10)$$

The d.c. equivalent circuit is then as shown in Fig. 3.8.

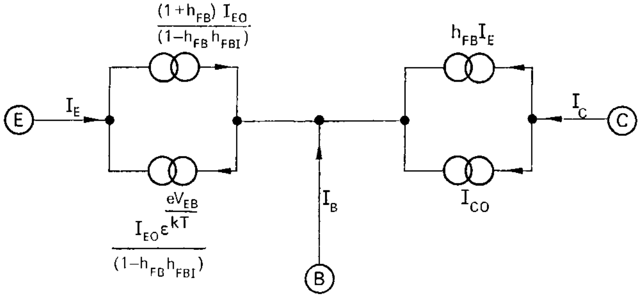

$\frac{(1+h_{FB})\,I_{EO}}{(1-h_{FB}h_{FBI})}$
$h_{FB}I_E$
E
I_E
I_C
C
I_B
I_{CO}
$\frac{I_{EO}\varepsilon^{\frac{eV_{EB}}{kT}}}{(1-h_{FB}h_{FBI})}$
B

FIG. 3.8. The d.c. equivalent circuit for a transistor.

The currents shown in Fig. 3.8 are in terms of h_{FB}, h_{FBI}, I_{EO}, and I_{CO} which, for a common-base *pnp* stage, are all negative, and if the moduli of these quantities are considered, the resulting equivalent circuit is as shown in Fig. 3.9.

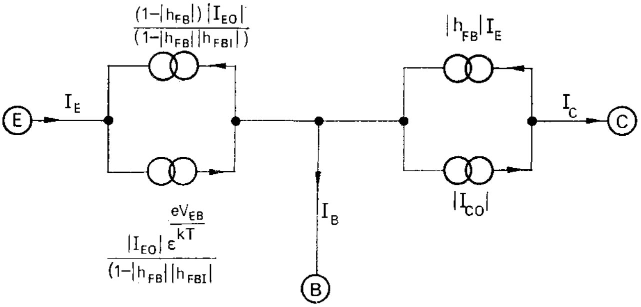

$\frac{(1-|h_{FB}|)\,|I_{EO}|}{(1-|h_{FB}|\,|h_{FBI}|)}$
$|h_{FB}|\,I_E$
E
I_E
I_C
C
I_B
$|I_{CO}|$
$\frac{|I_{EO}|\,\varepsilon^{\frac{eV_{EB}}{kT}}}{(1-|h_{FB}|\,|h_{FBI}|}$
B

FIG. 3.9. The d.c. equivalent circuit for a transistor showing true direction of current.

These Ebers–Moll relations given by eqns. (3.2.6) and (3.2.7) are useful in predicting the d.c. characteristics of transistors. As shown, they are for the common-base connection, but may easily be converted for use with the common emitter connection by means of the additional relations

$$I_E+I_C+I_B = 0 \quad \text{and} \quad V_{EB}+V_{BC}+V_{CE} = 0.$$

As an example of the use of the Ebers–Moll relations, consider what happens when both junctions are reverse biased, i.e. $\varepsilon^{(eV_{CB})/kT}$ and $\varepsilon^{(eV_{EB})/kT}$ negligible.

From eqn. (3.2.10),

$$I_E = \frac{(1+h_{FB})I_{EO}}{(1-h_{FB}h_{FBI})}.$$

In practice, h_{FB} is close to unity, and negative and h_{FBI} is considerably less but also negative. I_{EO} is also small. Therefore

$$I_E \doteq 0 \quad \text{and} \quad I_C = -I_{CO}.$$

A further example of the use of these equations is to determine the collector voltage to reduce the collector current to zero, i.e. from eqn. (3.2.9), putting $I_C = 0$,

$$h_{FB}I_E = I_{CO}(\varepsilon^{(eV_{CB})/kT}-1)$$

or

$$V_{CB} = \frac{kT}{e}\log_\varepsilon\left(1+\frac{h_{FB}I_E}{I_{CO}}\right).$$

Other examples of the use of the Ebers–Moll relations will no doubt occur to the reader.

Questions on Chapter 3

Q.1. Show from first principles that the potential barrier for a *pn* junction with zero bias is given by

$$V = \frac{kT}{e} \log_\varepsilon \left(\frac{N_D N_A}{N_i^2} \right),$$

where N_D and N_A are the impurity densities in the *n* and *p* sides of the junction and N_i the intrinsic carrier density.

Q.2. Derive the expression $I = I_S(\varepsilon^{(eV/kT)} - 1)$ relating the voltage and current for a *pn* junction.

Q.3. Assuming the diode equation and the relation between forward and reverse leakage currents and current gains, derive the Ebers–Moll relations for a transistor connected in the common-base configuration. Hence sketch the output characteristics.

Q.4. Repeat Q.3 for the common-emitter configuration.

CHAPTER 4

The Field Effect Transistor

ALTHOUGH the idea of a field effect transistor (FET) was first proposed in 1928, it was not until the late 1950's that manufacturing technology advanced sufficiently for these devices to become a practical proposition. Since they have not been covered in previous volumes in this series, a complete chapter is devoted to this topic, and a qualitative explanation of the operation is followed by a more theoretical approach. Typical circuits are discussed in Chapter 9.

4.1. Types of Field Effect Transistor

Field effect transistors are voltage-operated high-impedance devices that rely for their operation on the flow of majority carriers. In both junctions and metal oxide semiconductor (MOST) field effect transistors these majority carriers flow as a drift current along a channel, the conductivity of which is controlled by an externally applied voltage.

Junction field effect transistors were introduced in the late 1950's, and in this type the conducting channel, which may be either *n* or *p* type, lies between two reverse-biased *pn* junctions connected in parallel. The depletion regions associated with these junctions spread into the conducting channel and control its width as indicated in Figs. 4.1 and 4.2. Figure 4.1 shows the polarities and terminology used for an *n*-channel device, while Fig. 4.2 indicates the connection for a *p*-channel FET.

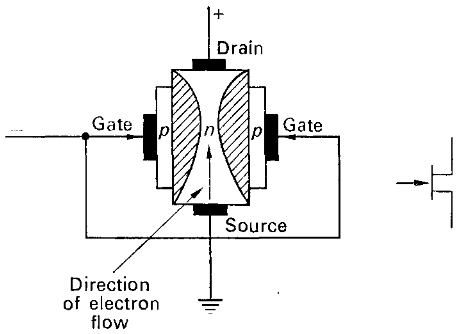


FIG. 4.1. An *n*-channel junction FET.

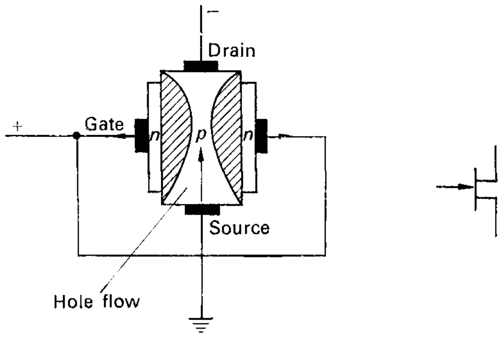


FIG. 4.2. A *p*-channel junction FET.

The charge carriers flow from the source to the drain, these electrodes being analogous to the cathode and anode of a vacuum triode while the conductivity of the channel is controlled by the voltage applied to the gate. This gate electrode, comparable to the grid of a vacuum triode, controls the width of the depletion region and hence the conductivity of the channel.

Metal oxide semiconductor transistors were first developed about 1962 and rely for their operation on the formation of a conducting layer along the interface between a semiconductor and an oxide layer produced beneath a control electrode. The construction of an n-channel MOST is illustrated by Fig. 4.3. The formation of the conducting path involves the inversion of the type of semiconductor and may be produced either by the field itself when the device

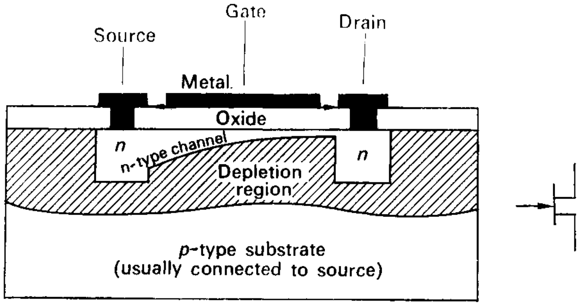


FIG. 4.3. An n-channel MOST.

is said to be operating under enhancement mode conditions or by light minority carrier doping in the depletion mode of operation.

4.2. The Junction Field Effect Transistor

As stated previously, the operation of this type depends on the width of a conducting channel being modified by the application of a voltage to two pn junctions on opposite sides of a semiconductor.

A typical test circuit for an n-channel device, showing the relevant variables, i.e. gate–source voltage V_{GS}, drain–source voltage V_{DS}, and drain current I_D, is shown in Fig. 4.4.

As the reverse voltage on the gate is increased, the width of the depletion layers will increase and that of the channel will decrease. In the limit when the gate–source voltage V_{GS} is sufficiently negative to reduce the width of the channel to zero by itself, no majority carriers will flow and the current is said to be "pinched off". The gate–source voltage V_{GS} under this condition is known as the pinch-off voltage V_P. To be more accurate, the pinch-off voltage

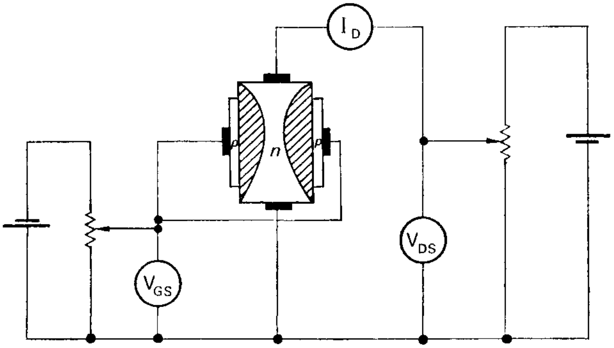


FIG. 4.4. Test circuit for junction FET.

is $V_{GS}+\phi$, where ϕ is the barrier potential barrier associated with an unbiased diode.

For any value of V_{GS} less negative than V_P, the voltage across the junctions will be negative and equal in magnitude to $V_{GS}+I_DR$, where I_DR is the voltage drop due to the channel itself. Since I_DR increases along the length of the channel and the gate electrodes are of finite length, the largest reverse voltage appears across that part of the junction nearest the drain. This means that the width of the depletion layer will increase as the drain end is approached and explains the variation in channel width shown in Fig. 4.4.

At a given value of V_{GS} (less than V_P) an increase of drain–source voltage V_{DS} from zero will initially give rise to a drain current I_D that is approximately proportional to the voltage V_{DS}. Under these conditions the device behaves as a resistance, and the operation is said to be ohmic. Further increase of V_{DS} will produce a state of affairs where $V_{GS}+I_DR$ is equal or greater than the pinch-off voltage, and the device is said to operate under "pinch-off" or saturation conditions. There must be a path for the drain current I_D to flow, but the path resistance increases with V_{DS} such that I_D remains very nearly constant. At some higher value of V_{DS} breakdown will occur between the drain and the gate.

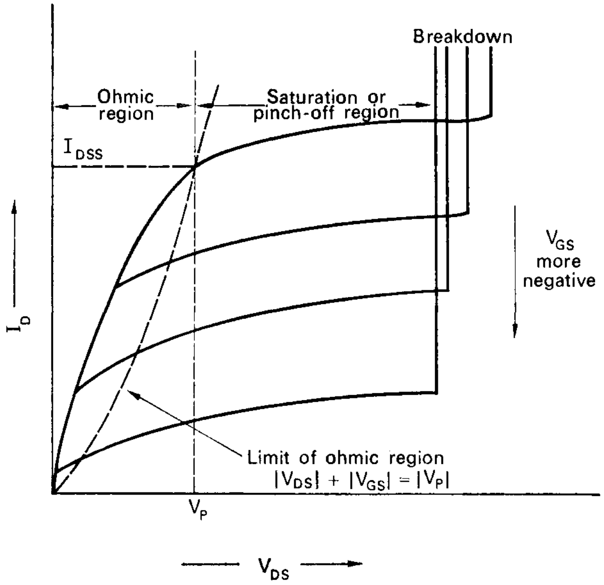


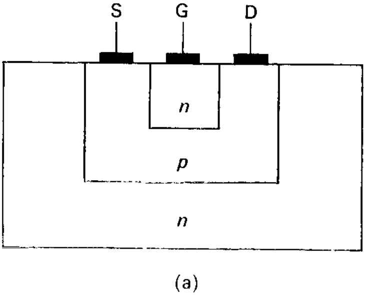
FIG. 4.5. Output characteristics of junction-type MOST.

A complete set of output characteristics is shown in Fig. 4.5. They are similar to those of a pentode, the saturation current being controlled by the gate–source voltage but dependent on the construction, i.e. open-channel resistance. Also on the curves is shown the limit of ohmic region given by the equation $|V_{DS}|+|V_{GS}| = |V_P|$ and I_{DSS}, the saturation current with zero gate–source voltage.

4.2.1. Construction of the Junction FET

The construction of the junction FET utilizes techniques that have been developed for bipolar transistors, e.g. diffusion and epitaxial processes.

Figure 4.6 (a) illustrates the construction of a *p*-channel diffused planar transistor. Starting with an *n*-type sample, successive diffusions of *p*- and *n*-type impurities give rise to a *npn* structure with a



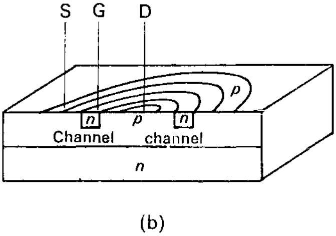


FIG. 4.6. Construction of junction-type FET.

p-type channel about 15 μm long connecting source and drain regions. Alternatively, a coaxial type of construction as shown in Fig. 4.6(b) may be produced. The impurity content of the gate region is normally considerably greater than that of the channel

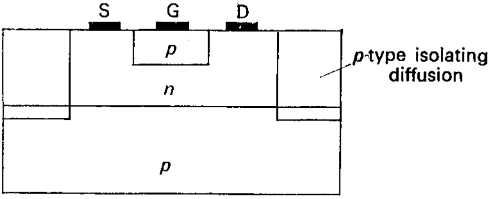


FIG. 4.7. Construction of epitaxial junction-type FET.

itself. Typically, a gate is 0·5 Ω-cm material compared with 4 Ω-cm for the channel.

Figure 4.7 shows how an *n*-channel FET may be constructed using an epitaxial technique. Initially, an *n*-type epitaxial layer is grown on a *p*-type substrate. Diffusion of *p*-type impurity follows to produce the gate region and end regions. The latter heavily doped regions extend into the *p*-type substrate and effectively isolate the field effect structure.

4.2.2. Theory of the junction FET

The operation of this type of FET obviously depends on the variation of the depletion layer thickness δ with gate voltage, and, initially, Poisson's equation will be used to derive an expression for this. As stated in the previous section, the impurity content of the gate region is considerably greater than that of the channel, and the junction is considered to be abrupt. The conditions appertaining to such a junction are shown in Fig. 4.8. Applying Poisson's equation to the *n*-type region,

$$\frac{d^2V}{dx^2} = -\frac{1}{\varepsilon_0\varepsilon_r}\times\text{space charge density},$$

where V is the voltage across the depletion region that extends into the *n*-type region and ε_0 and ε_r the free space and relative permittivities respectively.

In this depletion layer all the impurity atoms are ionized and the excess positive charges per unit volume must be equal to the number of impurity atoms per unit volume N_D,

i.e. $$\frac{d^2V}{dx^2} = -\frac{eN_D}{\varepsilon_0\varepsilon_r},$$

where e is the charge on an electron. Integrating with respect to x,

$$\frac{dV}{dx} = -\frac{eN_Dx}{\varepsilon_0\varepsilon_r}+C.$$

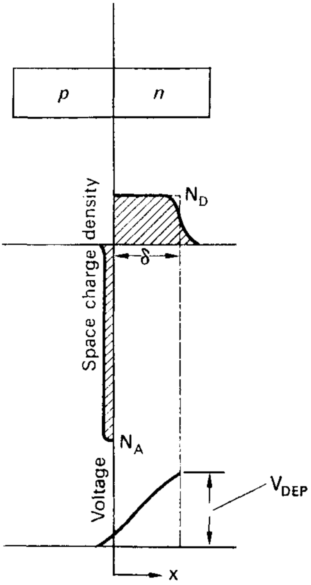


FIG. 4.8. Depletion layer voltage in a *pn* junction.

At $x = \delta$, the width of the depletion layer, $dV/dx = 0$, and hence the constant C is given by $C = eN_D\delta/\varepsilon_0\varepsilon_r$. Therefore

$$\frac{dV}{dx} = \frac{eN_D}{\varepsilon_0\varepsilon_r}(\delta - x).$$

Integrating again,

$$V = \frac{eN_D}{\varepsilon_0\varepsilon_r}\left(\delta x - \frac{x^2}{2}\right) + C_1.$$

With the original assumption that the impurity content of the *p*-type region is much greater than that of the *n*-type, i.e. virtually all the voltage appears across that part of the depletion layer that extends into the *n*-type region, at $x = 0$, $V = 0$, and, hence, $C_1 = 0$. Therefore

$$V = \frac{eN_Dx}{\varepsilon_0\varepsilon_r}\left(\delta - \frac{x}{2}\right).$$

The total depletion layer voltage V_{DEP} is the obtained by putting $x = \delta$,

i.e. $$V_{\mathrm{DEP}} = \frac{eN_D\delta^2}{2\varepsilon_0\varepsilon_r}$$

or $$\delta = \sqrt{\left(\frac{2\varepsilon_0\varepsilon_r V_{\mathrm{DEP}}}{eN_D}\right)}.$$

Consider now the *n*-channel junction FET shown in Fig. 4.9. Electrons will flow from the source to the drain until the voltage across the *pn* junction at the gate is such that the width of the depletion layer is $D/2$, where D is the channel width,

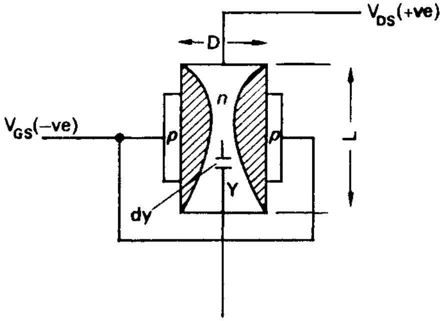


FIG. 4.9. Analysis of junction-type FET.

i.e. $$\frac{D}{2} = \sqrt{\left[\frac{2\varepsilon_0\varepsilon_r(V_{GS}+\phi)}{eN_D}\right]},$$

where ϕ is the barrier potential, i.e. the voltage across the junction when $V_{GS} = 0$. This is about 0·3 V for germanium and 0·7 V for silicon. Therefore

$$V_{GS}+\phi = \frac{eN_D D^2}{8\varepsilon_0\varepsilon_r}.$$

This is known as the pinch-off voltage V_P. Hence,

$$V_P = \frac{eN_D D^2}{8\varepsilon_0\varepsilon_r}.$$

When the gate voltage is less than the pinch-off voltage V_P, the voltage across the junction at any point Y is given by

$$V_{\text{DEP}} = V_{GS}+\phi+I_D R,$$

where I_D is the drain current and R is the resistance between the source and the point Y.

For a small distance dy the incremental resistance is given by

$$dR = \frac{\varrho\, dy}{(D-2\delta)B},$$

where B is the transverse dimension of the semiconductor and ϱ is the resistivity. Therefore

$$R = \int_0^y \frac{\varrho\, dy}{(D-2\delta)B}$$

and $$V_{\text{DEP}} = V_{GS}+\phi+I_D \int_0^y \frac{\varrho\, dy}{(D-2\delta)B}.$$

Differentiating with respect to y,

$$\frac{dV_{\text{DEP}}}{dy} = \frac{I_D\varrho}{(D-2\delta)B}.$$

But $$\delta = \sqrt{\left(\frac{2\varepsilon_0\varepsilon_r V_{\text{DEP}}}{eN_D}\right)} = \frac{D}{2}\sqrt{\left(\frac{V_{\text{DEP}}}{V_P}\right)}$$

since $$V_P = \frac{eN_D D^2}{8\varepsilon_0\varepsilon_r}.$$

Hence, substituting for δ,

$$D\left[1-\sqrt{\left(\frac{V_{\text{DEP}}}{V_P}\right)}\right] dV_{\text{DEP}} = \frac{I_D\varrho}{B}\, dy.$$

Integrating over the length L of the semiconductor,

$$\int_{V_1}^{V_2} D\left[1-\sqrt{\left(\frac{V_{\text{DEP}}}{V_P}\right)}\right] dV_{\text{DEP}} = \int_0^L \frac{I_D\varrho}{B}\,dy.$$

V_1 and V_2 correspond to the depletion layer voltages near the source and the drain. Near the source the depletion layer voltage may be regarded as being due to the gate–source voltage alone, while near the drain it may be regarded as being the sum of the magnitudes of V_{GS} and V_{DS};

i.e. $\quad V_1 = V_{GS}+\phi \quad$ and $\quad V_2 = V_{GS}+V_{DS}+\phi,$

where ϕ is the barrier potential associated with an unbiased diode.

Therefore $$D\left(V_{\text{DEP}}-\frac{2}{3}\,\frac{V_{\text{DEP}}^{3/2}}{V_P^{1/2}}\right)_{V_{GS}+\phi}^{V_{GS}+V_{DS}+\phi} = \frac{I_D\varrho L}{B}$$

or $$I_D = \frac{BD}{\varrho L}\left\{V_{DS}-\frac{2}{3\sqrt{(V_P)}}\left[(V_{GS}+V_{DS}+\phi)^{3/2}-(V_{GS}+\phi)^{3/2}\right]\right\}.$$

When $V_{GS} = 0$, saturation occurs when $V_{DS}+\phi$ is equal to the pinch-off voltage V_P. The value of the saturation current (I_{DSS}) under this condition is given by

$$I_{DSS} = \frac{BD}{\varrho L}\left[(V_P-\phi)-\frac{2}{3\sqrt{(V_P)}}(V_P^{3/2}-\phi^{3/2})\right].$$

Hence $$I_{DSS} = \frac{BD}{\varrho L}\left\{\frac{V_P}{3}-\phi\left[1-\frac{2}{3}\sqrt{\left(\frac{\phi}{V_P}\right)}\right]\right\}.$$

To a first approximation, the second term may be ignored, giving

$$I_{DSS} = \frac{BDV_P}{3\varrho L}.$$

The drain current may now be written in terms of I_{DSS},

$$\text{i.e.}\quad I_D = I_{DSS}\left\{3\left(\frac{V_{DS}}{V_P}\right)-2\left[\left(\frac{V_{GS}+V_{DS}+\phi}{V_P}\right)^{3/2}-\left(\frac{V_{GS}+\phi}{V_P}\right)^{3/2}\right]\right\}.$$

Saturation occurs at any other value of V_{GS} when $V_{DS}+V_{GS}+\phi = V_P$.

Hence the saturation drain current I_{DS} for any value of V_{GS} is given by

$$\begin{aligned} I_{DS} &= I_{DSS}\left\{3\left(\frac{V_P-V_{GS}-\phi}{V_P}\right)-2\left[1-\left(\frac{V_{GS}+\phi}{V_P}\right)^{3/2}\right]\right\} \\ &= I_{DSS}\left[1-3\left(\frac{V_{GS}+\phi}{V_P}\right)+2\left(\frac{V_{GS}+\phi}{V_P}\right)^{3/2}\right]. \end{aligned}$$

Alternatively, the drain current at the the onset of saturation for any value of V_{DS} may be determined by putting $V_{GS} = V_P-\phi-V_{DS}$ in the expression for I_D,

$$\begin{aligned} \text{i.e.}\quad I_D &= I_{DSS}\left\{3\left(\frac{V_{DS}}{V_P}\right)-2\left[1-\left(\frac{V_P-V_{DS}}{V_P}\right)^{3/2}\right]\right\} \\ &= I_{DSS}\left[3\left(\frac{V_{DS}}{V_P}\right)-2+2\left(1-\frac{V_{DS}}{V_P}\right)^{3/2}\right] \\ &= I_{DSS}\left\{3\left(\frac{V_{DS}}{V_P}\right)-2+2\left[1-\frac{3}{2}\left(\frac{V_{DS}}{V_P}\right)+\frac{3}{8}\left(\frac{V_{DS}}{V_P}\right)^2\dots\right]\right\} \\ &= I_{DSS}\left[\frac{3}{4}\left(\frac{V_{DS}}{V_P}\right)^2+\dots\right]. \end{aligned}$$

(V_{DS}/V_P) is equal to unity for $V_{GS} = 0$, but gets progressively smaller as pinch off is approached. To a first approximation $I_D = \frac{3}{4}I_{DSS}\,(V_{DS}/V_P)^2$, but to allow for higher order terms the expression for the limit of the ohmic region is often given as

$$I_D = I_{DSS}\left(\frac{V_{DS}}{V_P}\right)^2.$$

When considering small variations, the mutual conductance is an important parameter. This is analogous with vacuum valves and is defined as the change in drain saturation current with gate voltage when the drain voltage is constant.

In the expression

$$I_{DS} = I_{DSS}\left[1-3\left(\frac{V_{GS}+\phi}{V_P}\right)+2\left(\frac{V_{GS}+\phi}{V_P}\right)^{3/2}\right],$$

V_{GS} is actually a negative quantity. Therefore

$$\begin{aligned} g_m &= -\frac{\partial I_{DS}}{\partial V_{GS}} \\ &= -I_{DSS}\left[\frac{3}{V_P}+\frac{2}{V_P^{3/2}}\times\frac{3}{2}\,(V_{GS}+\phi)^{1/2}\right] \\ &= \frac{3I_{DSS}}{V_P}\left[1-\sqrt{\left(\frac{V_{GS}+\phi}{V_P}\right)}\right]. \end{aligned}$$

This has a maximum value g_{m_0} when V_{GS} is zero, i.e. $g_{m_0} = 3I_{DSS}/V_P$ to a first approximation.

At any other value of V_{GS},

$$g_m = g_{m_0}\left[1-\sqrt{\left(\frac{V_{GS}+\phi}{V_P}\right)}\right].$$

The fact that g_m varies with the gate voltage gives a non-linear transfer characteristic that is very nearly square law.

Similar expressions may be derived for linear graded junctions, but the initial expression for the thickness of the depletion layer is modified. This leads to an expression for the drain current given by

$$I_D = I_{DSS}\left\{\frac{5}{2}\left(\frac{V_{DS}}{V_P}\right)-\frac{3}{2}\left[\left(\frac{V_{GS}+V_{DS}+\phi}{V_P}\right)^{5/3}-\left(\frac{V_{GS}+\phi}{V_P}\right)^{5/3}\right]\right\}.$$

4.3. The Insulated Gate Field Effect Transistor (Enhancement Type)

This is one of several variations of FET in which an insulating layer of silicon dioxide separates the gate or control electrode from a surface-conducting channel. The d.c. operation may be illustrated with reference to Fig. 4.10.

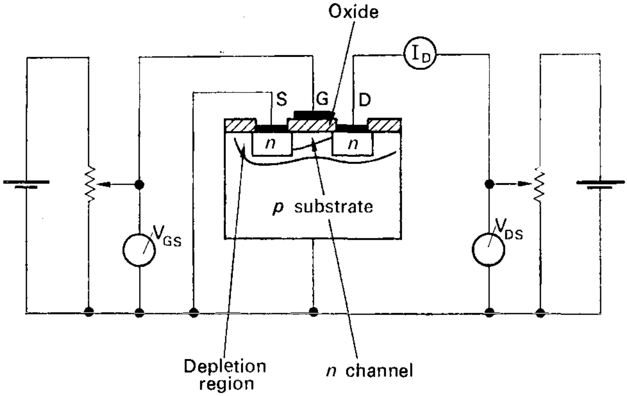


FIG. 4.10. Test circuit of a MOST (enhancement mode).

With zero voltage on the gate, the *p* region extends between the heavily doped source and drain regions effectively isolating these regions. The drain current I_D under this condition is very nearly zero for all values of drain–source voltage V_{DS} as long as it is less than the break-down value.

If the gate is positive with respect to the substrate, the gate electrode, oxide layer, and substrate form a capacitor, and a negative charge is induced at the semiconductor–oxide interface. This "ielectron enhancement" results in a conducting path being produced between the source and drain regions and initially, as V_{DS} is increased with V_{GS} maintained constant, there is a corresponding increase in drain current. As the drain–source voltage V_{DS} is increased further, the voltage drop between the source and drain must be considered. This gives rise to a variation in potential across the "plates" of the capacitor as shown in Fig. 4.11. Since the width of the conducting channel is a function of the gate substrate voltage, the channel width decreases as the drain contact is approached. A consequent limitation in the drain current is produced, and further increase in V_{DS} only leads to an increase in channel resistance such that the drain current is very nearly constant. This condition corresponds to "pinch off" in the junction FET.

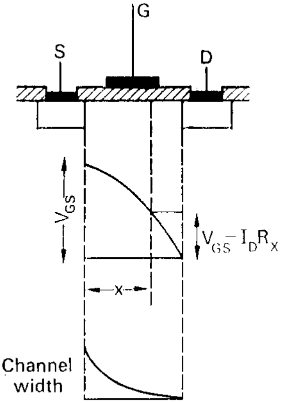


FIG. 4.11. Variation in channel width for a MOST.

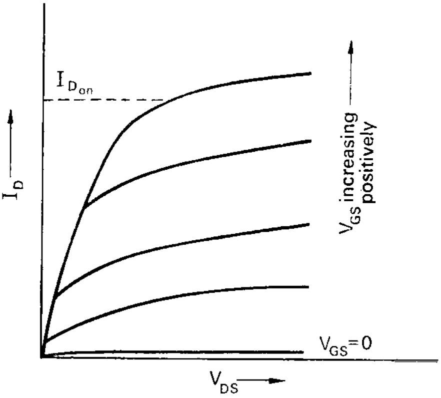


FIG. 4.12. Output characteristics for an enhancement-type MOST.

A complete set of output characteristics for an *n*-channel MOST is shown in Fig. 4.12, positive increments of V_{GS} giving non-linear increases in drain saturation current. These characteristics assume the substrate is connected to the source, and under such conditions the drain saturation current I_{DSS} is very nearly zero. Any potential applied between substrate and source will result in a bodily shift of the characteristics, and indicates a possible application to mixing circuits.

4.3.1. Construction of an enhancement-type MOST

The construction of this type is illustrated by Fig. 4.13. Starting with a relatively high resistivity *p*-type substrate, low-resistivity, *n*-type source and drain regions are first produced by diffusion. The oxide coating is the next step followed by etch-cutting of windows to allow evaporation of aluminium for the source, gate, and drain electrodes.

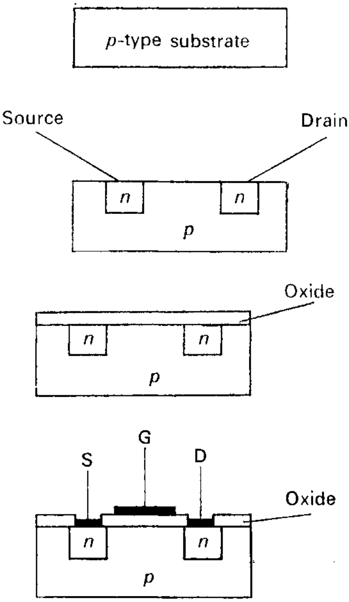


FIG. 4.13. Construction of an enhancement-type MOST.

4.3.2. Theory of the enhancement-type MOST

A theoretical treatment of the insulated gate transistor was first published by S. R. Hofstein and F. P. Heinman in the September 1963 issue of the *Proceedings of the IEEE*, and readers are recommended to read this original paper.

If an *n*-channel device is connected as shown in Fig. 4.14, the voltage across the oxide layer at any point Y is given by $V_{GS}-V_Y$, where V_{GS} is the gate–source voltage and V_Y the voltage at Y due to the drain–source voltage V_{DS}. This induces a negative surface charge density q on the oxide–semiconductor interface,

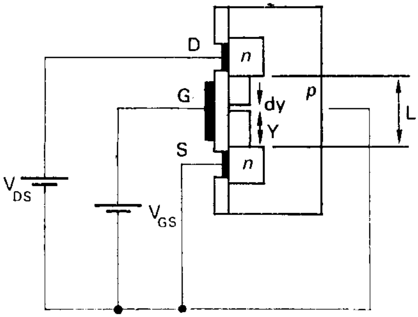


FIG. 4.14. Analysis of an enhancement-type MOST.

i.e. $$q = \varepsilon_0\varepsilon_r E,$$

where ε_0 is the permittivity of free space, ε_r is the relative permittivity of the oxide, and E is the field strength.

If, as is normally true, the channel is much narrower than the oxide,

$$q = \varepsilon_0\varepsilon_r\left(\frac{V_{GS}-V_Y}{d}\right),$$

where d is the thickness of the oxide layer.

Some of the electrons comprising this charge are free to move but some are trapped, mainly in incomplete bonds, at the semiconductor–oxide interface;

i.e. $$q = \text{mobile charge} + \text{immobile charge}.$$

The latter may be regarded as being due to a voltage V_P, where V_P is comparable to the pinch-off voltage in a junction FET. In this context, V_P may be defined as the maximum positive voltage that corresponds to zero mobile carriers. Therefore

$$\text{mobile charge density} = \frac{\varepsilon_0\varepsilon_r(V_{GS}-V_Y-V_P)}{d} \quad \text{coulombs m}^2.$$

This equation is obviously only valid for V_{GS} greater than or equal to V_Y+V_P.

The drain current I_D flowing is then given by

$$I_D = \frac{\varepsilon_0\varepsilon_r(V_{GS}-V_Y-V_P)}{d} Bv_E,$$

where B is the transverse dimension of the device and v_E is the drift velocity of electrons.

But
$$\begin{aligned} v_E &= \mu_E E \\ &= \mu_E \frac{dV_Y}{dy}, \end{aligned}$$

where μ_E is the mobility, assumed constant.

Therefore
$$I_D = \frac{\varepsilon_0\varepsilon_r(V_{GS}-V_Y-V_P)}{d} B\mu_E \frac{dV_Y}{dy},$$

and, rearranging,
$$\frac{I_D d}{\mu_E\varepsilon_0\varepsilon_r B}\,dy = (V_{GS}-V_Y-V_P)\,dV_Y\,.$$

Integrating over the length of the channel

$$\frac{I_D d}{\mu_E\varepsilon_0\varepsilon_r B}\int_0^L dy = \int_0^{V_{DS}} (V_{GS}-V_Y-V_P)\,dV_Y$$

or
$$\frac{I_D\,dL}{\mu_E\varepsilon_0\varepsilon_r B} = (V_{GS}-V_P)V_{DS}-\frac{1}{2}V_{DS}^2$$

or
$$I_D = \frac{\mu\varepsilon_0\varepsilon_r B}{dL}\left[(V_{GS}-V_P)V_{DS}-\frac{1}{2}V_{DS}^2\right].$$

This analysis ignores the physical resistance of the source and drain regions which, in fact, modifies the limits of integration.

The limit of the ohmic region may then be determined by putting $V_{GS} = V_P + V_{DS}$,

i.e. $$I_D = \frac{\mu\varepsilon_0\varepsilon_r B}{2\,dL} V_{DS}.$$

In this type of MOST, I_{DSS} is very nearly zero and usually $I_{D_{\mathrm{ON}}}$ is taken as a reference value. This is the current flowing when the gate is biased for maximum conduction.

Putting $V_{GS} = V_{GS_{\max}}$ and $V_{DS} = V_{GS_{\max}} - V_P$ in the expression for I_D,

$$I_{D_{\mathrm{ON}}} = \frac{\mu\varepsilon_0\varepsilon_r B}{dL}\left[(V_{GS_{\max}} - V_P)^2 - \frac{1}{2}(V_{GS_{\max}} - V_P)^2\right]$$

$$= \frac{\mu\varepsilon_0\varepsilon_r B}{2\,dL}(V_{GS_{\max}} - V_P)^2.$$

Therefore $$I_D = \frac{2I_{D_{\mathrm{ON}}}}{(V_{GS_{\max}} - V_P)^2}\left[(V_{GS} - V_P)V_{DS} - \frac{1}{2}V_{DS}^2\right].$$

Putting $V_{DS} = V_{GS} - V_P$ enables the drain saturation current for any value of V_{GS} to be evaluated,

i.e. $$I_{DS} = \frac{2I_{D_{\mathrm{ON}}}}{(V_{GS_{\max}} - V_P)^2}\left[(V_{GS} - V_P)^2 - \frac{1}{2}(V_{GS} - V_P)^2\right]$$

$$= I_{D_{\mathrm{ON}}}\left(\frac{V_{GS} - V_P}{V_{GS_{\max}} - V_P}\right)^2.$$

Obviously, when $V_{GS} = V_{GS_{\max}}$, this reduces to $I_{D_{\mathrm{ON}}}$.

The mutual conductance, again defined as the change in drain saturation current with gate voltage, when the drain voltage is constant, is given by

$$g_m = \frac{\partial I_{DS}}{\partial V_{GS}}$$

$$= \frac{I_{D_{\mathrm{ON}}}}{(V_{GS_{\max}} - V_P)^2} \times 2(V_{GS} - V_P).$$

This has a maximum value of g_{m_0} when $V_{GS} = V_{GS_{\max}}$,

i.e. $$g_{m_0} = \frac{2I_{D_{ON}}}{(V_{GS_{\max}} - V_P)}.$$

At any other value of V_{GS},

$$g_m = g_{m_0}\left(\frac{V_{GS} - V_P}{V_{GS_{\max}} - V_P}\right).$$

This expression shows that the mutual conductance is directly proportional to the gate voltage V_{GS}.

4.4. The Insulated Gate Field Effect Transistor (Depletion Type)

This type of MOST transistor differs from the previous one in that during manufacture a conducting channel is produced by diffusion in the substrate between the source and the drain. The resulting channel allows an appreciable current to flow even when the gate voltage is zero, and as the gate is taken progressively more negative there is a corresponding reduction in current.

Complete d.c. connections for a MOST device operating under depletion conditions are shown in Fig. 4.15. The drain current that flows when the gate substrate voltage is zero, increases approxi-

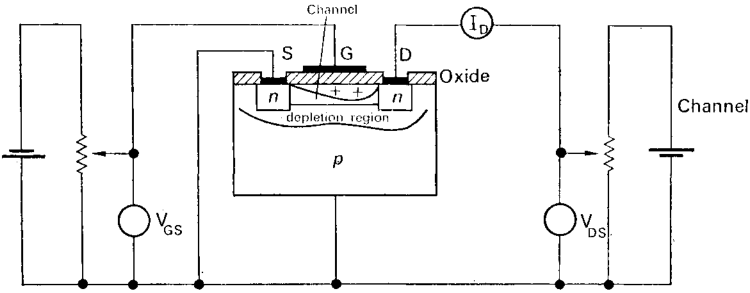


FIG. 4.15. Test circuit of a MOST (depletion mode).

mately linearly with drain–source voltage for low values of V_{DS}, but soon reaches its saturation value. This saturation or "pinch-off" effect is due, as in the enhancement type, to the variation in voltage along the channel itself. The voltage between gate and substrate becomes progressively more negative as the drain is approached, depleting the channel of carriers and causing a variation in channel width as shown in Fig. 4.15.

Similar characteristics are obtained for negative potentials applied to the gate, but obviously a lower saturation current is reached at a lower value of drain–source voltage. A complete set of characteristics is shown in Fig. 4.16, and, with the assumption that the substrate and source are connected, the drain saturation current for zero gate voltage I_{DSS} is relatively high. As in the enhancement-mode type equal increments of V_{GS} produce unequal increments of I_{DS}.

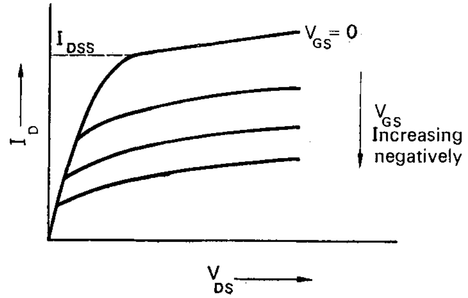


Fig. 4.16. Output characteristics of a MOST (depletion mode).

4.4.1. Construction of a depletion-type MOST

The construction of this type of device follows the same pattern as the enhancement mode type with the addition of an extra stage of diffusion. This takes place before oxidization and a medium resistive *n*-type channel is produced in the substrate between the high resistivity source and drain regions, as indicated in Fig. 4.15.

4.4.2. Theory of the depletion-type MOST

As stated previously, the presence of free electrons just below the oxide layer in an *n*-channel device gives rise to a conducting channel when the gate voltage is zero. This corresponds to a surface charge density q_c.

$$\text{Then mobile charge density } = q_c+\frac{\varepsilon_0\varepsilon_r(V_{GS}-V_Y-V_P)}{d}$$

$$= \frac{\varepsilon_0\varepsilon_r}{d}(V_{GS}-V_Y-V_P+V_C),$$

where $V_C = \dfrac{q_c d}{\varepsilon_0\varepsilon_r}$.

Hence, as in the previous section,

$$I_D = \frac{\mu\varepsilon_0\varepsilon_r B}{dL}\left[(V_{GS}-V_P+V_C)V_{DS}-\frac{1}{2}V_{DS}^2\right].$$

In fact for this type of device, maximum current corresponds to $V_{GS} = 0$, and in practice V_{GS} is made negative to produce a positive charge at the interface and deplete the channel of carriers. Maximum current I_{DSS} then corresponds to $V_{GS} = 0$ and $V_C = V_P+V_{DS}$,

i.e. $$I_{DSS} = \frac{\mu\varepsilon_0\varepsilon_r B}{2\,dL}(V_C-V_P)^2.$$

Hence at a given positive value of V_{DS} and a given negative value of V_{GS} less than pinch off, the current is given by

$$I_D = \frac{2I_{DSS}}{(V_C-V_P)^2}\left[(V_C-V_P-V_{GS})V_{DS}-\frac{1}{2}V_{DS}^2\right].$$

The limit of the ohmic region may then be determined by putting $V_C = V_P+V_{GS}+V_{DS}$,

i.e. $$I_D = \frac{2I_{DSS}}{(V_C-V_P)^2}\left[V_{DS}^2-\frac{1}{2}V_{DS}^2\right]$$

$$= I_{DSS}\left(\frac{V_{DS}}{V_C-V_P}\right)^2.$$

The drain saturation current at any value of V_{GS} may be evaluated by substituting $V_{DS} = V_C - V_P - V_{GS}$ in the expression for the drain saturation current,

i.e. $$I_{DS} = \frac{2I_{DSS}}{(V_C - V_P)^2}\left[(V_C - V_P - V_{GS})^2 - \frac{1}{2}(V_C - V_P - V_{GS})^2\right]$$

$$= I_{DSS}\left(\frac{V_C - V_P - V_{GS}}{V_C - V_P}\right)^2$$

$$= I_{DSS}\left[1 - \left(\frac{V_{GS}}{V_C - V_P}\right)\right]^2.$$

The output characteristics for this device are then as shown in Fig. 4.16.

Finally, in this section the mutual conductance may be determined,

i.e. $$g_m = -\frac{\partial I_{DS}}{\partial V_{GS}}$$

$$= -I_{DSS}2\left[1 - \left(\frac{V_{GS}}{V_C - V_P}\right)\right]\left(-\frac{1}{V_C - V_P}\right)$$

$$= \frac{2I_{DSS}}{(V_C - V_P)}\left[1 - \frac{V_{GS}}{V_C - V_P}\right].$$

This has a maximum value g_{m_0} when V_{GS} is zero,

i.e. $$g_{m_0} = \frac{2I_{DSS}}{V_C - V_P}.$$

At any other value of V_{GS}

$$g_m = g_{m_0}\left[1 - \frac{V_{GS}}{V_C - V_P}\right].$$

This shows that g_m is directly proportional to the gate voltage.

4.5. The Insulated Gate Field Effect Transistor (Enhancement Depletion Type)

Although the MOST devices described in the previous section are limited to a maximum gate voltage of zero, they may be modified such that the gate may be made positive with respect to the source without exceeding the current rating. When this occurs, the device will operate under depletion mode conditions for negative gate–source voltages and enhancement mode conditions for positive gate–source voltages. Devices designed as such have the advantages that no bias components are required, and signals to be amplified may be connected directly to the gate.

A typical set of output characteristics for such devices is shown in Fig. 4.17. The drain saturation current (I_{DSS}) for zero gate–source voltage is something less than half the maximum allowable current due to the non-linear relation between drain saturation current and gate–source voltage.

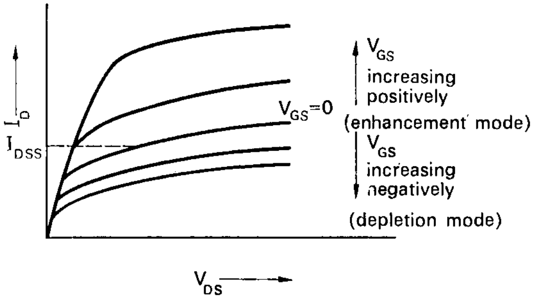


FIG. 4.17. Output characteristics of a MOST (enchancement/depletion mode).

The construction follows the same pattern as the depletion type, but the amount of impurity diffused into the substrate material for the *n*-type channel is controlled so that both depletion and enhancement modes of operation are possible.

A theoretical treatment will obviously follow the same pattern as the depletion type MOST with the stipulations that V_{GS} can go positive and I_{DSS} is not the maximum current.

4.6. Equivalent Circuit of a Field Effect Transistor

As in the case of vacuum valves, a FET is a voltage-operated device with a very high input impedance. The input current may be ignored for the majority of purposes and the drain current may be regarded as a function of gate and drain voltages,

i.e. $$I_D = f(V_{GS}, V_{DS}).$$

Using the total differential theorem,

$$\delta I_D = \frac{\partial I_D}{\partial V_{GS}} \delta V_{GS} + \frac{\partial I_D}{\partial V_{DS}} \delta V_{DS}$$

$$= g_m \delta V_{GS} + \frac{\delta V_{DS}}{r_{DS}}, \qquad (4.6.1)$$

where $g_m = \dfrac{\partial I_D}{\partial V_{GS}}$ is the mutual conductance

and $r_{DS} = \dfrac{\partial V_{DS}}{\partial I_D}$ is the drain resistance.

Equation (4.6.1) may be represented by the equivalent circuit shown in Fig. 4.18.

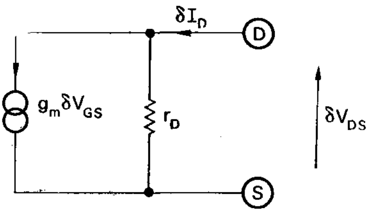


FIG. 4.18. Low-frequency equivalent circuit of a FET.

As shown, the diagram represents small changes in direct values, but it is equally valid for small-signal low-frequency alternating currents.

At higher frequencies, device capacitances have to be included. The input capacitance C_{GS} is that of the reverse biased *pn* junction for the junction-type FET, and that of the natural capacitor formed by the gate, oxide and channel for the MOST. The output capacitance C_{DS} is not usually as important as the input capacitance, due to the lower resistances associated with it, but the drain–gate capacitance must be allowed for in the case of MOST devices. This is due to the inherent constructional difficulty of ensuring that the gate electrode does not spread over the drain region as indicated in Fig. 4.19. The oxide layer between them again provides the dielectric for a natural capacitor. An equivalent circuit valid for high-frequency calculations is then as shown in Fig. 4.20.

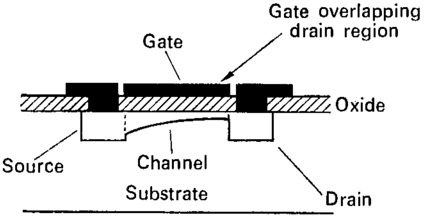


FIG. 4.19. Drain–gate capacitance of a MOST.

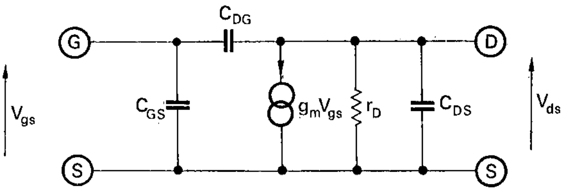


FIG. 4.20. High-frequency equivalent circuit of a FET.

Questions on Chapter 4

Q.1. Discuss qualitatively the construction and characteristics of FETs. Both junction and insulated-gate types should be considered.

Q.2. Show that for an insulated-gate field-effect transistor the drain current in the ohmic region is given by

$$I_D = \frac{2I_{DSS}}{(V_C - V_P)^2}\left[(V_C - V_P - V_{GS})V_{DS} - \frac{1}{2}V_{DS}^2\right],$$

where the symbols have the usual meaning.

Q.3. Show that the pinch-off voltage for a junction-type FET is given by

$$V_P = \frac{eN_D D^2}{8\varepsilon_0\varepsilon_r}.$$

What is the comparable effect in an insulated-gate FET?

Q.4. The mutual conductance of a FET depends on the gate source voltage, Explain quantitatively how this statement applies to (a) a junction-type FET. (b) an insulated-gate FET.

Q.5. Derive an equivalent circuit for a FET that is valid at low frequencies. How is this modified for high-frequency operation?

CHAPTER 5

Small-signal Equivalent Circuits of Bipolar Transistors

OVER the years a series of small-signal equivalent circuits for bipolar transistors have been suggested, and of these it appears that four are well established. Two of these, the "h" and "y", are black-box equivalent circuits and are derived directly from input and output conditions, while the other two, the "T" and the hybrid-π, bear some relation to the physical construction of the device. A different subdivision is that "h" and "T" equivalent circuits are in general confined to low frequencies, while "y" and hybrid-π are more useful at high frequencies.

5.1. The "h" Equivalent Circuit

This circuit has been used extensively in previous volumes and is derived from Fig. 5.1 (a) and shown in Fig. 5.1 (b). Since any three of the four variables V_1, I_1, V_2, I_2 will ensure a unique value of the fourth, they are not unrelated variables, and any one may be regarded as a function of any other two. For this circuit, V_1 and I_2 are assumed to be functions of I_1 and V_2,

i.e. $$V_1 = f(I_1, V_2) \quad \text{and} \quad I_2 = F(I_1, V_2).$$

Using the total differential theorem,

$$\delta V_1 = \frac{\partial V_1}{\partial I_1}\delta I_1 + \frac{\partial V_1}{\partial V_2}\delta V_2 \quad \text{and} \quad \delta I_2 = \frac{\partial I_2}{\partial I_1}\delta I_1 + \frac{\partial I_2}{\partial V_2}\delta V_2 .$$

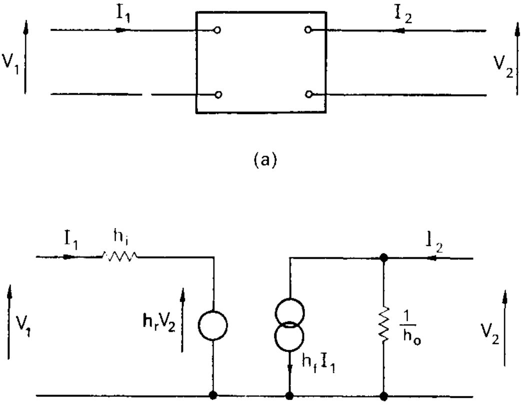


FIG. 5.1. The "h" equivalent circuit.

$\frac{\partial V_1}{\partial I_1} = h_i$ is the slope of the input characteristic or the a.c. input impedance with output short circuited.

$\frac{\partial V_1}{\partial V_2} = h_r$ is the slope of the feedback characteristics or the a.c. voltage feedback ratio with input open circuited.

$\frac{\partial I_2}{\partial I_1} = h_f$ is the slope of the forward characteristic or the a.c. forward current gain with output short circuited.

$\frac{\partial I_2}{\partial V_2} = h_o$ is the slope of the output characteristic or the output admittance with input open circuited.

It does not matter whether *pnp* or *npn* transistors are considered since the slopes of the characteristics are of the same sign. Also it does not matter whether small changes in d.c. or small-signal a.c. is considered, and hence these relations may be reduced to

$$V_1 = h_iI_1 + h_rV_2 \quad \text{and} \quad I_2 = h_fI_1 + h_oV_2.$$

These equations are directly represented by the equivalent circuit shown in Fig. 5.1 (b).

With a load admittance Y_l the equivalent circuit is as shown in Fig. 5.2, and

$$V_2 = -\frac{h_f I_1}{Y_l + h_o}.$$

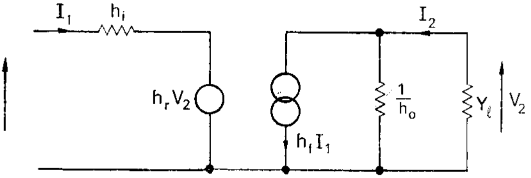


FIG. 5.2. The "h" equivalent circuit including load admittance.

Therefore $$V_1 = h_i I_1 + h_r V_2$$
$$= h_i I_1 + h_r\left(-\frac{h_f I_1}{Y_l + h_o}\right).$$

Input impedance, $$Z_{\text{in}} = \frac{V_1}{I_1} = h_i - \frac{h_r h_f}{Y_l + h_o}. \qquad (5.1.1)$$

Also $$I_2 = h_f I_1 + h_o\left(-\frac{h_f I_1}{Y_l + h_o}\right)$$

or current gain $$\frac{I_2}{I_1} = h_f - \frac{h_o h_f}{Y_l + h_o}$$
$$= -\frac{h_f Y_l}{Y_l + h_o}. \qquad (5.1.2)$$

Also, voltage gain $= \frac{V_2}{V_1}$

$$= -\frac{h_f I_1}{(Y_l + h_o)} \times \frac{1}{I_1\left(h_i - \frac{h_r h_f}{Y_l + h_o}\right)}$$
$$= -\frac{h_f}{h_i(Y_l + h_o) - h_r h_f}. \qquad (5.1.3)$$

Further, assuming a source impedance of Z_S and applying a voltage V to the output as shown in Fig. 5.3,

$$I_1 = -\frac{h_r V}{h_i + Z_S}.$$

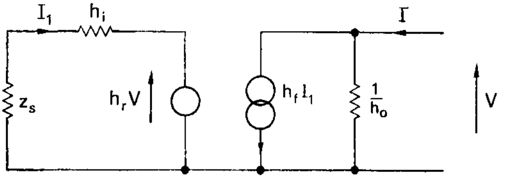


FIG. 5.3. Output admittance using h parameters.

Therefore
$$I = Vh_o - \frac{h_f h_r V}{h_i + Z_S},$$

i.e. output admittance
$$Y_{\text{out}} = \frac{I}{V}$$
$$= h_o - \frac{h_r h_f}{h_i + Z_S}. \qquad (5.1.4)$$

Assuming all the h parameters are purely resistive (valid over the usual range of calculations) and the source and load impedances are also purely resistive, the matching resistances may be derived from eqns. (5.1.1) and (5.1.4), i.e. eliminating Z_S between these equations,

$$Y_l = h_o - \frac{h_r h_f}{h_i + h_i - \dfrac{h_r h_f}{Y_l + h_o}}$$

or
$$(Y_l - h_o)\,[2h_i(Y_l + h_o) - h_r h_f] = -h_r h_f (Y_l + h_o),$$

i.e.
$$2h_i(Y_l^2 - h_o^2) - h_r h_f (Y_l - h_o) = -h_r h_f (Y_l + h_o)$$

or
$$2h_i(Y_l^2 - h_o^2) = -2h_r h_f h_o,$$

i.e.

$$\begin{aligned} Y_l &= \sqrt{\left(h_o^2 - \frac{h_r h_f h_o}{h_i}\right)} \\ &= \sqrt{\left(\frac{h_o}{h_i}\right)}\sqrt{[(h_o h_i - h_r h_f)]} \\ &= \sqrt{\left(\frac{h_o h}{h_i}\right)}, \end{aligned}$$

where

$$h = h_o h_i - h_r h_f. \tag{5.1.5}$$

Therefore

$$\begin{aligned} Z_S &= h_i - \frac{h_r h_f}{\sqrt{\left(\frac{h_o h}{h_i}\right)} + h_o} \\ &= \frac{\sqrt{(h_o h h_i)} + h_o h_i - h_r h_f}{\sqrt{\left(\frac{h_o h}{h_i}\right)} + h_o} \\ &= \frac{\sqrt{(h_o h h_i)} + h}{\sqrt{\left(\frac{h_o h}{h_i}\right)} + h_o} \\ &= \frac{\sqrt{h}\,[\sqrt{(h_o h_i)} + \sqrt{h}]}{\sqrt{\frac{h_o}{h_i}}\,[\sqrt{h} + \sqrt{(h_o h_i)}]} \\ &= \sqrt{\left(\frac{h h_i}{h_o}\right)}. \end{aligned} \tag{5.1.6}$$

The a.c. power gain may also be derived. For pure resistive source impedance R_S and pure resistance load impedance

$$R_l = \frac{1}{G_l},$$

$$\begin{aligned} \text{power gain} &= \frac{P_{\text{out}}}{P_{\text{in}}} \\ &= \frac{V_2^2 G_l}{V_1^2 / R_{\text{in}}} \\ &= \frac{h_f^2}{[h_i(G_l + h_o) - h_r h_f]^2}\, G_l \left[h_i - \frac{h_r h_f}{G_l + h_o}\right] \\ &= \frac{h_f^2 G_l}{[h_i(G_l + h_o) - h_r h_f]\,(G_l + h_o)}. \end{aligned} \tag{5.1.7}$$

The maximum available power gain occurs when the transistor is correctly matched,

i.e. maximum available power gain

$$= \frac{h_f^2\sqrt{\left(\frac{h_o h}{h_i}\right)}}{\left\{h_i\left[\sqrt{\left(\frac{h_o h}{h_i}\right)}+h_o\right]-h_r h_f\right\}\left[\sqrt{\left(\frac{h_o h}{h_i}\right)}+h_o\right]}$$

$$= \frac{h_f^2\sqrt{\left(\frac{h_o h}{h_i}\right)}}{[\sqrt{(h_i h_o h)}+h]\left[\sqrt{\left(\frac{h_o h}{h_i}\right)}+h_o\right]}$$

$$= \frac{h_f^2}{[\sqrt{(h_o h h_i)}+h]\left[1+\sqrt{\left(\frac{h_o h_i}{h}\right)}\right]}$$

$$= \frac{h_f^2}{\sqrt{(h_o h h_i)}+h+h_o h_i+\sqrt{(h_o h_i h)}}$$

$$= \frac{h_f^2}{[\sqrt{h}+\sqrt{(h_o h_i)}]^2}. \tag{5.1.8}$$

These relations are valid for common-base, common-emitter and common-collector circuits, suitable subscripts being used in each case. The relation between these parameters have already been derived (see Volume 2, section 7.3) and are summarized in Table 1.

The comparison between parameters brings out some interesting results for the common collector.

From eqn. (5.1.1), $Z_{\text{in}} = h_{ie}+\frac{(1+h_{fe})}{Y_l+h_{oe}}$

$$\simeq h_{fe}Z_l.$$

From eqn. (5.1.2), current gain

$$\simeq h_{fe}.$$

TABLE 1

Common base	Common emitter	Common collector
$h_{ib}=\dfrac{h_{ie}}{1+h_{fe}}$	h_{ie}	$h_{ic}=h_{ie}$
$h_{rb}=\dfrac{h_{ie}h_{oe}}{1+h_{fe}}$	h_{re}	$h_{rc}=1$
$h_{fb}=-\dfrac{h_{fe}}{1+h_{fe}}$	h_{fe}	$h_{fc}=-(1+h_{fe})$
$h_{ob}=\dfrac{h_{oe}}{1+h_{fe}}$	h_{oe}	$h_{oc}=h_{oe}$

From eqn. (5.1.3), voltage gain

$$\triangleq 1.$$

From eqn. (5.1.4), output admittance

$$\triangleq h_{oe}+\frac{(1+h_{fe})}{h_{ie}+Z_s}\triangleq\frac{h_{fe}}{Z_s},$$

i.e. output impedance

$$\triangleq\frac{Z_s}{h_{fe}}.$$

From eqn. (5.1.8), maximum available power gain

$$=\frac{(1+h_{fe})^2}{(\sqrt{[h_{ie}h_{oe}+(1+h_{fe})]}+\sqrt{(h_{ie}h_{oe})})^2}$$
$$\triangleq h_{fe},$$

since h_{fe} is the predominant term.

Although not usually used at high frequencies, some writers have modified the low-frequency real values of h_i, h_f, and h_o by the inclusion of quadrature components. These include an input capacitance in parallel with h_i, an output capacitance in parallel with h_o, and a modification of the current gain h_f. The input capacitance is composed of components due to the emitter–base depletion layer capacitances and the much larger diffusion capacitances. Even when combined, however, their reactances are usually negligible compared with h_i. The output capacitance due to the collector depletion layer must be included, and the variation in h_f is as given in previous volumes,

i.e.
$$h_{fb} = \frac{h_{fbo}}{1+jf/f_\alpha}$$

where h_{fbo} is the common base current gain at low frequencies and f_α is the α cut-off frequency, and

$$h_{fe} = \frac{h_{feo}}{1+jf/f_\beta}$$

where h_{feo} is the common emitter current gain at low frequencies and f_β is the β cut-off frequency.

A high-frequency small-signal equivalent circuit for the common-base connection would then be as shown in Fig. 5.4. For the common-emitter connection a similar circuit could be drawn, the output capacitance in the two circuits being related in a similar manner to the output admittances,

i.e.
$$h_{oe} = h_{ob}(1+h_{fe}),$$

and
$$C_{oe} = C_{ob}(1+h_{fe}).$$

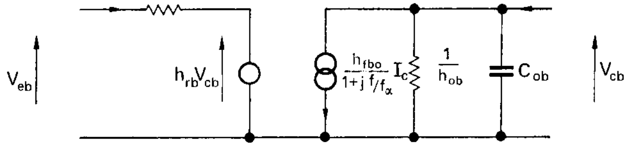


FIG. 5.4. High-frequency "h" equivalent circuit.

5.2. The "y" Equivalent Circuit

As with the "h" equivalent circuit, this representation of a transistor has been used in the previous volume and is redrawn in Fig. 5.5.

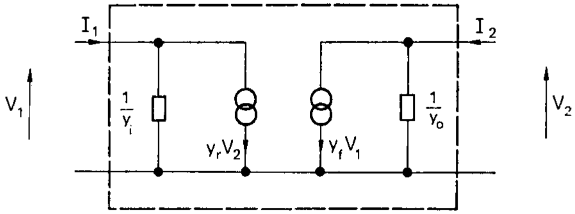


FIG. 5.5. The "y" equivalent circuit.

In this case the input and output currents are assumed to be functions of V_1 and V_2,

i.e. $I_1 = f(V_1, V_2)$ and $I_2 = F(V_1, V_2)$,

using the total differential theorem

$$\delta I_1 = \frac{\partial I_1}{\partial V_1}\delta V_1 + \frac{\partial I_1}{\partial V_2}\delta V_2 \quad \text{and} \quad \delta I_2 = \frac{\partial I_2}{\partial V_1}\delta V_1 + \frac{\partial I_2}{\partial V_2}\delta V_2,$$

where

$\frac{\partial I_1}{\partial V_1} = y_i$ is the input admittance with the output short circuited.

$\frac{\partial I_1}{\partial V_2} = y_r$ is the reverse transfer admittance, i.e. the ratio of the current flowing in the input circuit to the voltage applied to the output circuit with the input short circuited.

$\frac{\partial I_2}{\partial V_2} = y_o$ is the output admittance with the input short circuited.

$\frac{\partial I_2}{\partial V_1} = y_f$ is the forward transfer admittance, i.e. the ratio of output current to input voltage with the output short circuited.

These are basically a.c. parameters and are easily measured on admittance bridges. The above expressions may then be rewritten

$$I_1 = y_i V_1 + y_r V_2 \quad \text{and} \quad I_2 = y_f V_1 + y_o V_2$$

where y_i, y_r, y_f and y_o are in general complex.

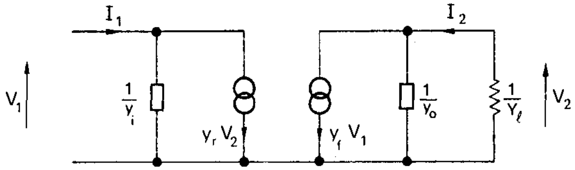


FIG. 5.6. The "y" equivalent circuit including load admittance.

Including a load admittance Y_l, the equivalent circuit for an amplifier using y parameters is as shown in Fig. 5.6,

i.e. $$I_2 = -Y_l V_2.$$

Hence $$-Y_l V_2 = y_f V_1 + y_o V_2$$

or voltage gain $$\frac{V_2}{V_1} = -\frac{y_f}{Y_l + y_o}. \qquad (5.2.1)$$

Also, current gain, $$\frac{I_2}{I_1} = \frac{y_f V_1 + y_o V_2}{y_i V_1 + y_r V_2}$$

$$= \frac{y_f - (y_o y_f)/(Y_l + y_o)}{y_i - (y_r y_f)/(Y_l + y_o)}$$

$$= \frac{y_f Y_l}{y_i(Y_l + y_o) - y_r y_f}. \qquad (5.2.2)$$

Eliminating V_2 from the expression, $I_1 = y_i V_1 + y_r V_2$,

$$I_1 = y_i V_1 - \frac{y_r y_f V_1}{Y_l + y_o},$$

i.e. input admittance $$Y_{\text{in}} = \frac{I_1}{V_1} = y_i - \frac{y_r y_f}{Y_l + y_o}. \qquad (5.2.3)$$

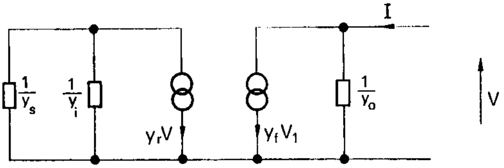


FIG. 5.7. Output admittance using *y* parameters.

Further, assuming a source admittance y_s and applying a voltage V to the output as shown in Fig. 5.7, the voltage across the input is given by

$$-\frac{y_r V}{y_s+y_i}.$$

Hence the current I is given by

$$I = y_o V - \frac{y_f y_r}{y_s+y_i} V,$$

and output admittance

$$Y_{\text{out}} = \frac{I}{V} = y_o - \frac{y_f y_r}{y_s+y_i}. \tag{5.2.4}$$

Remembering that all the y parameters are complex, the input and output admittances are given by

$$g_{\text{in}}+jb_{\text{in}} = (g_i+jb_i) - \frac{(g_r+jb_r)(g_f+jb_f)}{(g_l+jb_l)+(g_o+jb_o)}$$

and $$g_{\text{out}}+jb_{\text{out}} = (g_o+jb_o) - \frac{(g_r+jb_r)(g_f+jb_f)}{(g_s+jb_s)+(g_i+jb_i)}.$$

For correct matching,

$$g_s = g_{\text{in}}, \qquad b_s = -b_{\text{in}},$$
$$g_l = g_{\text{out}}, \qquad b_l = -b_{\text{out}}.$$

Hence under these conditions,

$$g_s - jb_s = (g_i+jb_i) - \frac{(g_r+jb_r)(g_f+jb_f)}{(g_l+g_o)+j(b_l+b_o)},$$

$$g_l - jb_l = (g_o+jb_o) - \frac{(g_r+jb_r)(g_f+jb_f)}{(g_s+g_i)+j(b_s+b_i)}.$$

Therefore
$$[(g_s-g_i)-j(b_s+b_i)]\,[(g_l+g_o)+j(b_l+b_o)]$$
$$= -(g_r g_f-b_r b_f)-j(g_r b_f+g_f b_r) \qquad (5.2.5)$$

and
$$[(g_l-g_o)-j(b_l+b_o)]\,[(g_s+g_i)+j(b_s+b_i)]$$
$$= -(g_r g_f-b_r b_f)-j(g_r b_f+g_f b_r). \qquad (5.2.6)$$

Subtracting,

$$2g_s g_o-2g_i g_l-2jg_l(b_s+b_i)+2jg_s(b_l+b_o) = 0,$$

equating real terms, $g_s g_o = g_i g_l$,

equating quadrature terms, $g_l(b_s+b_i) = g_s(b_l+b_o)$,

from these expressions $g_o(b_s+b_i) = g_i(b_l+b_o)$;

adding eqns. (5.2.5) and (5.2.6),

$$2g_s g_l-2g_i g_o+2(b_s+b_i)(b_l+b_o)-2jg_o(b_s+b_i)-2jg_i(b_l+b_o)$$
$$= -2(g_r g_f-b_r b_f)-2j(g_r b_f+g_f b_r),$$

equating quadrature terms, $g_o(b_s+b_i)+g_i(b_l+b_o) = (g_r b_f+g_f b_r)$.

Since $g_o(b_s+b_i) = g_i(b_l+b_o)$, $b_s = \left(\dfrac{g_r b_f+g_f b_r}{2g_o}\right)-b_i$

and $b_l = \left(\dfrac{g_r b_f+g_f b_r}{2g_i}\right)-b_o$,

equating real terms,

$$g_s g_l-g_i g_o+(b_s+b_i)(b_l+b_o) = -(g_r g_f-b_r b_f).$$

Substituting from the above relations

$$g_s^2\frac{g_o}{g_i}-g_i g_o+\frac{(g_r b_f+g_f b_r)^2}{4g_i g_o} = -(g_r g_f-b_r b_f)$$

or
$$g_s = \sqrt{\left[g_i^2-\frac{g_i}{g_o}(g_r g_f-b_r b_f)-\left(\frac{g_r b_f+g_f b_r}{2g_o}\right)^2\right]}$$

and
$$g_l = \sqrt{\left[g_o^2-\frac{g_o}{g_i}(g_r g_f-b_r b_f)-\left(\frac{g_r b_f+g_f b_r}{2g_i}\right)^2\right]}.$$

The power gain for any terminating impedance is given by

$$\frac{P_{\text{out}}}{P_{\text{in}}} = \left|\frac{V_2}{V_1}\right|^2 \frac{g_l}{g_{\text{in}}}$$

$$= \frac{|y_f|^2}{|Y_l + y_o|^2} \times \frac{g_l}{\text{Re}\left[y_i - \dfrac{y_r y_f}{Y_l + y_o}\right]},$$

where $\text{Re}\left[y_i - \dfrac{y_r y_f}{Y_l + y_o}\right]$ is the real part of the input impedance.

Under matched conditions, the maximum available power gain is achieved,

i.e. maximum available power gain

$$= \frac{(g_f + b_f)^2}{(g_l + g_o)^2 + (b_l + b_o)^2} \times \frac{g_l}{g_{\text{in}}},$$

where $$g_l = \sqrt{\left[g_o^2 - \frac{g_o}{g_i}(g_r g_f - b_r b_f) - \left(\frac{g_r b_f + g_f b_r}{2g_i}\right)^2\right]}$$

$$\frac{g_l}{g_{\text{in}}} = \frac{g_o}{g_i}$$

and $$(b_l + b_o) = \left(\frac{g_r b_f + g_f b_r}{2g_l}\right)$$

Therefore maximum available power gain

$$= \frac{(g_f + b_f)^2 \dfrac{g_o}{g_i}}{\begin{array}{l} 2g_o^2 - \dfrac{g_o}{g_i}(g_r g_f - b_r b_f) - \left(\dfrac{g_r b_f + g_f b_r}{2g_l}\right)^2 \\ \quad + 2g_o \sqrt{\left[g_o^2 - \dfrac{g_o}{g_i}(g_r g_f - b_r b_f)\left(\dfrac{g_r b_f + g_f b_r}{2g_1}\right)^2\right]} + \left(\dfrac{g_r b_f + g_f b_r}{2g_i}\right)^2 \end{array}}$$

$$= \frac{(g_f + b_f)^2 \dfrac{g_o}{g_i}}{2g_o^2 - \dfrac{g_o}{g_i}(g_r g_f - b_r b_f) + 2g_o \sqrt{\left[g_o^2 - \dfrac{g_o}{g_i}(g_r g_f - b_r b_f) - \left(\dfrac{g_r b_f + g_f b_r}{2g_i}\right)^2\right]}}$$

$$= \frac{(g_f + b_f)^2}{2g_o g_i - (g_r g_f - b_r b_f) + \sqrt{[4g_i^2 g_o^2 - 4g_o g_i(g_r g_f - b_r b_f) - (g_r b_f + g_f b_r)^2]}}$$

As in the case of hybrid parameters, these relations are valid for common-base, common-emitter, and common-collector circuits, suitable subscripts being used in each case. The relations between these may easily be derived as shown below.

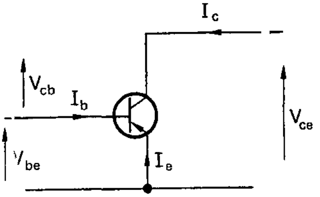


FIG. 5.8. Currents and voltages for a common-emitter stage.

Considering the common-emitter circuit shown in Fig. 5.8,

$$I_b = V_{be}y_{ie}+V_{ce}y_{re},$$

$$I_c = V_{be}y_{fe}+V_{ce}y_{oe}.$$

But
$$I_e+I_b+I_c = 0$$

and
$$V_{ce} = V_{cb}+V_{be};$$

hence
$$\begin{aligned}I_c &= (-V_{eb})y_{fe}+(V_{cb}-V_{eb})y_{oe}\\ &= V_{eb}(-y_{fe}-y_{oe})+V_{cb}y_{oe},\end{aligned}$$

and by comparison with $I_c = V_{eb}y_{fb}+V_{cb}y_{ob}$,

$$y_{fb} = -(y_{fe}+y_{oe}) \quad \text{and} \quad y_{ob} = y_{oe}.$$

Also
$$\begin{aligned}-I_e &= I_b+I_c\\ &= (-V_{eb})y_{ie}+(V_{cb}-V_{eb})y_{re}\\ &\quad +(-V_{eb})y_{fe}+(V_{cb}-V_{eb})y_{oe}\end{aligned}$$

or
$$I_e = V_{eb}(y_{ie}+y_{re}+y_{oe}+y_{fe})+V_{cb}(-y_{re}-y_{oe}),$$

and by comparison with $I_e = V_{eb}y_{ib}+V_{cb}y_{rb}$,

$$y_{ib} = y_{ie}+y_{re}+y_{fe}+y_{oe} \quad \text{and} \quad y_{rb} = -(y_{re}+y_{oe}).$$

TABLE 2

Common base	Common emitter	Common collector
$y_{ib} = (y_{ie}+y_{re}+y_{fe}+y_{oe})$	y_{ie}	$y_{ic} = y_{ie}$
$y_{rb} = -(y_{re}+y_{oe})$	y_{re}	$y_{rc} = -y_{re}$
$y_{fb} = -(y_{fe}+y_{oe})$	y_{fe}	$y_{fc} = -(y_{fe}+y_{ie})$
$y_{ob} = y_{oe}$	y_{oe}	$y_{oc} = y_{oe}+y_{re}$

In a similar manner, relations between common-emitter and common-collector parameters may be derived. Table 2 is complete.

For low-frequency purposes, when reactive terms may be ignored, the y parameters may easily be related to the h parameters,

i.e. $$V_1 = h_i I_1 + h_r V_2$$

and $$I_2 = h_f I_1 + h_o V_2.$$

The first of these equations may be rearranged to give

$$I_1 = \frac{1}{h_i} V_1 - \frac{h_r}{h_i} V_2.$$

Hence, by comparison with $I_1 = y_i V_1 + y_r V_2$,

$$y_i = \frac{1}{h_i} \quad \text{and} \quad y_r = -\frac{h_r}{h_i}.$$

Also $$I_2 = h_f\left(\frac{V_1}{h_i} - \frac{h_r V_2}{h_i}\right) + h_o V_2$$

$$= \frac{h_f}{h_i} V_1 + \left(h_o - \frac{h_r h_f}{h_i}\right) V_2,$$

and by comparison with $I_2 = y_f V_1 + y_o V_2$,

$$y_f = \frac{h_f}{h_i} \quad \text{and} \quad y_o = h_o - \frac{h_r h_f}{h_i}.$$

As stated previously, these y parameters are very useful for high frequencies, particularly when the frequency range is such that the variation in parameters with frequency may be ignored.

As an example of the use of the y parameters, consider the unilateralized amplifier shown in Fig. 5.9. The bias components have been omitted from this circuit and the relevant theory, covered in previous volumes, is assumed.

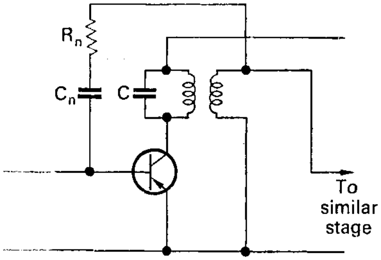


FIG. 5.9. Unilateralized amplifier.

The amplifier has a bandwidth of 200 kHz centred on 10 MHz, the loss-free transformer is tightly coupled, and the admittance parameters y_{ie}, y_{oe}, and y_{re} are 2 mS, in parallel with 50 pF, 50 μS in parallel with 2 pF and $50\ \underline{/270^\circ}$ μS respectively.

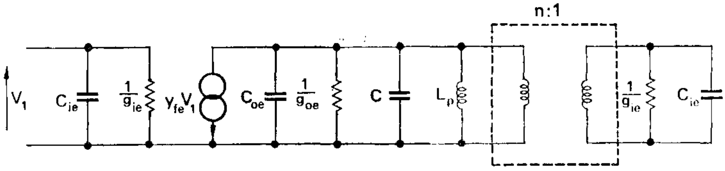


FIG. 5.10. Equivalent circuit of Fig. 5.9.

The unilateralizing components C_n and R_n will exactly balance the internal feedback through the device, represented by $y_{re}V_{ce}$, and the equivalent circuit of the stage is as shown in Fig. 5.10.

Assuming perfect unilaterization, the output resistance of the stage is $1/g_{oe} = 10^6/50 = 20\ \text{k}\Omega$, and for this to match the input resistance of $1/g_{ie} = 10^3/2 = 500\ \Omega$, the turns ratio is given by

$$n = \sqrt{\left(\frac{20\,000}{500}\right)} = \sqrt{40} = 6{\cdot}33.$$

The output circuit is then represented by Fig. 5.11 and for a bandwidth of 200 kHz centred on 10 MHz

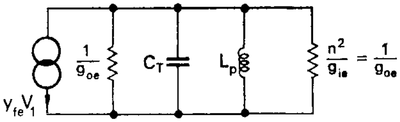


FIG. 5.11. Output circuit of Fig. 5.9.

$$Q = \frac{f_0}{\text{Bandwidth}} = \frac{R_{\text{eff}}}{\omega L_p},$$

where L_p is the primary inductance, and R_{eff} is the effective valve of the damping resistance (10 kΩ),

i.e. $$50 = \frac{10^4}{2\pi \times 10^7 \times L_p},$$

i.e. $$L_p = \frac{10^4}{\pi \times 10^9}\ \text{H} = 3{\cdot}2\ \mu\text{H}.$$

The total tuning capacitance is given by

$$\omega^2 = \frac{1}{LC_T}$$

or $$C_T = \frac{1}{4\pi^2 \times 10^{14} \times 3{\cdot}2 \times 10^{-5}}\ \text{F}$$
$$= 79{\cdot}5\ \text{pF};$$

2 pF of this is supplied by the output admittance y_{oe} and $\frac{50}{40} =$ 1·25 pF is reflected across the transformer. Hence the required tuning capacitance is 79·5 − 3·3 = 76 pF.

Under the unilateralized condition, an input V_1 gives rise to a current $y_{fe}V_1$ at the collector and $y_{fe}V_1Z_c/6{\cdot}33$ across the secondary of the transformer. The internal feedback represented by $-y_{re}y_{fe}V_1Z_c$ must be balanced by that due to the unilateralizing network. Allowing for the necessary inversion by the transformer, this gives

$$-y_{re}y_{fe}V_1Z_c = \left(\frac{\dfrac{y_{fe}V_1Z_c}{6{\cdot}33} - V_1}{Z\underline{/\phi}}\right)$$

or
$$y_{re}y_{fe}V_1Z_c = -\frac{y_{fe}V_1Z_c}{6{\cdot}33Z\underline{/\phi}} \text{ very nearly,}$$

where $Z\underline{/\phi}$ is the impedance of the unilateralizing network and Z_c the impedance in the output circuit,

i.e.
$$\begin{aligned} Z\underline{/\phi} &= -\frac{1}{6{\cdot}33y_{re}} \\ &= -\frac{1}{6{\cdot}33\times 50\times 10^{-6}\underline{/270^\circ}} \\ &= -3160\underline{/-90^\circ}\,\Omega. \end{aligned}$$

Hence the unilateralizing network comprises a capacitor whose value is given by

$$\begin{aligned} C_n &= \frac{1}{2\pi\times 10^7\times 3160}\text{ F} \\ &= 0{\cdot}5\text{ F}. \end{aligned}$$

Within the limits of calculation the unilateralizing resistance is zero.

5.3. The "T" Equivalent Circuit

This equivalent circuit was widely used in the early days of transistors. It is still used extensively, but has tended to be replaced by the "h" equivalent circuit or a simplified version of the hybrid-π.

With correct operating potentials, a transistor comprises a forward-biased emitter junction between a low resistivity emitter and a relatively high resistivity base, and a reverse-biased collector junction between the base and low resistivity collector. The resistances of the emitter and collector regions are very low, but account must be taken of the appreciable base resistance, and this may be shown as a resistance $r_{bb'}$, where b is the external base connection and b' an imaginary internal base connection. Figure 5.12 shows the complete low-frequency circuit with r_e as the resistance of the forward-biased emitter diode and r_c the resistance of the reverse-biased collector diode. Current and voltage generators complete the circuit, the current generator αI_e representing the proportion of the emitter current I_e which appears at the collector, while the voltage generator represents feedback between the collector and emitter circuits. The latter is necessary since an increase in collector voltage (i.e. more negative) leads to an increase in depletion layer thickness and decrease in base width, which implies a lower emitter current for a given collector current.

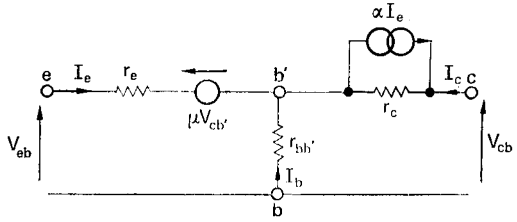


FIG. 5.12. Grounded-base "T" equivalent circuit.

The factor μ, however, is not an independent variable since it is related to the other parameters r_e, r_c, and α. Consider a small increase δV_{eb} in emitter voltage giving rise to an increase in emitter current δI_e, and an increase in collector current of very nearly $-\delta I_e$.

Then
$$\delta V_{cb'} = (\alpha\,\delta I_e + \delta I_c)\, r_c$$
$$= -\delta I_e(1-\alpha)\, r_c.$$

But $$\delta I_e = \frac{\delta V_{eb'} - \mu\,\delta V_{cb'}}{r_e}$$

since $$\delta V_{eb} \simeq \delta V_{eb'}.$$

Therefore $$\delta V_{cb'} = -\frac{(\delta V_{eb'} - \mu\delta V_{cb'})}{r_e}(1-\alpha)\,r_c,$$

i.e. $$\frac{r_e}{(1-\alpha)\,r_c} = -\left(\frac{\delta V_{eb'}}{\delta V_{cb'}} - \mu\right).$$

Allowing for the difference in sign between $V_{eb'}$ and $V_{cb'}$,

$$\mu = -\frac{\delta V_{eb'}}{\delta V_{cb'}}.$$

Hence $$\mu = \frac{r_e}{2(1-\alpha)\,r_c}.$$

The circuit is not normally used in this form and is usually transformed into the circuit shown in Fig. 5.13. The resistance $r_{b'}$ is an additional resistance to allow for the feedback generator $\mu V_{cb'}$. Its value may be estimated by considering the input circuits of Figs. 5.12 and 5.13,

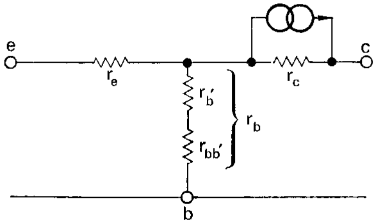


FIG. 5.13. Simplified grounded-base "T" equivalent circuit.

i.e. $$V_{eb} = I_e r_e + \mu V_{cb'} + (I_e + I_c)\,r_{bb'} = I_e r_e + (I_e + I_c)\,(r_{b'} + r_{bb'})$$

or $$\mu V_{cb'} = (I_e + I_b)\,r_{b'}.$$

But $$V_{bc'} = (I_e+I_c)\,r_c \qquad \text{(assuming } \alpha = 1\text{)},$$

hence $$r_{b'} = \mu r_c$$
$$= \frac{r_e}{2(1-\alpha)}.$$

The resistance $r_{b'}$ is known as the intrinsic base resistance compared with $r_{bb'}$ which is termed the extrinsic base resistance. Combined, these resistances are designated r_b. As indicated, the equivalent circuit shown in Fig. 5.12 corresponds to the common-base connection. It may easily be reorientated to give circuits for the common-emitter and common-collector connections as shown in Figs. 5.14 and 5.15.

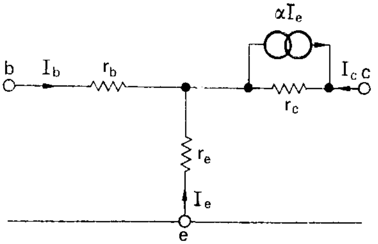


FIG. 5.14. Simplified grounded-emitter "T" equivalent circuit.

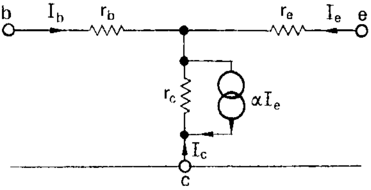


FIG. 5.15. Simplified grounded-collector "T" equivalent circuit.

The current generator in the common-emitter circuit is often stated in terms of the base current. If the volt drop across r_c is considered,

i.e. $$V_{r_c} = I_c r_c + \alpha I_e r_c$$
$$= I_c r_c - \alpha(I_c+I_b)\,r_c \qquad (\text{since } I_e+I_b+I_c = 0)$$
$$= I_c r_c(1-\alpha) - \alpha I_b r_c$$
$$= r_c(1-\alpha)\left[I_c - \left(\frac{\alpha}{1-\alpha}\right) I_b\right]$$
$$= r_c(1-\alpha)\,[I_c - \beta I_b],$$

where $$\beta = \frac{\alpha}{1-\alpha}.$$

This equation shows that if the impedance r_c is replaced by $r_c(1-\alpha)$, the current generator may be replaced by $-\beta I_b$. Hence the equivalent circuit may be re-drawn as shown in Fig. 5.16.

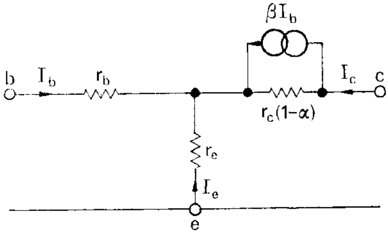


FIG. 5.16. Modified grounded-emitter "T" equivalent circuit.

The analysis of circuits using the "T" equivalent circuit gives different expressions for the three modes of connection, not just a change of subscript as with "h" and "y" equivalent circuits.

Consider the common-emitter circuit with a load R_l as shown in Fig. 5.17.

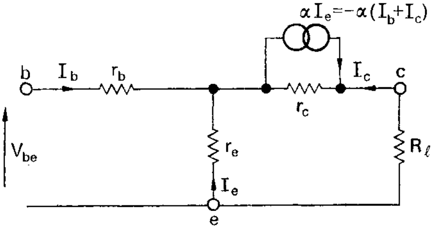


FIG. 5.17. Common-emitter equivalent circuit including load resistance.

The equations for the circuit are

$$V_{be} = I_b r_b + (I_b + I_c) r_e \tag{5.3.1}$$

and

$$0 = I_c(R_l + r_c + r_e) + I_b r_e - \alpha(I_b + I_c) r_c. \tag{5.3.2}$$

The current gain may be determined directly from eqn. (5.3.2),

i.e. $$\text{current gain} = \frac{I_c}{I_b}$$

$$= \frac{\alpha r_c - r_e}{R_l + r_e + r_c(1-\alpha)}$$

$$= \frac{\alpha}{(1-\alpha)+\dfrac{R_l}{r_c}} \quad \text{since } r_e \ll r_c.$$

Eliminating I_c from eqn. (5.31),

$$V_{be} = I_b(r_b + r_e) + \frac{\alpha I_b r_e}{(1-\alpha)+(R_l/r_c)}.$$

Therefore $$Z_{\text{in}} = \frac{V_{be}}{I_b}$$

$$= r_b + r_e + \frac{\alpha r_e}{(1-\alpha)+(R_l/r_c)}$$

$$= r_b + \frac{r_e}{1-\alpha} \quad \text{if} \quad r_c \gg R_l.$$

$$\text{Voltage gain} = -\frac{I_c R_l}{I_b R_{\text{in}}}$$

$$= -\frac{\alpha}{(1-\alpha)} \times \frac{R_l}{r_b + [r_e/(1-\alpha)]}$$

$$= -\frac{\alpha R_l}{r_e + r_b(1-\alpha)}.$$

The power gain with a pure resistance load is given by

$$\text{Power gain} = \text{current gain} \times \text{voltage gain},$$

$$= \frac{\alpha^2 R_l}{(1-\alpha)\,[r_e + r_b(1-\alpha)]}.$$

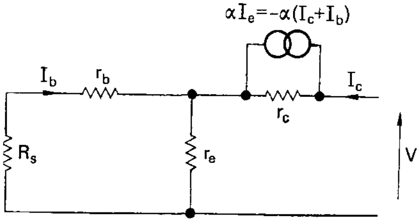


FIG. 5.18. Output impedance using T parameters.

To find the output impedance a voltage V may be applied to the output with the input generator replaced by its output impedance (Fig. 5.18). Then

$$V = I_c(r_c+r_e)-\alpha(I_c+I_b)r_c+I_br_e$$

and

$$0 = I_b(R_s+r_b+r_e)+I_cr_e\,.$$

Eliminating I_b gives

$$V = I_c[r_e+r_c(1-\alpha)]+\frac{(\alpha r_c+r_e)r_e}{R_s+r_b+r_e}\,.$$

Since

$$r_e \ll r_c\,,$$

$$Z_{\text{out}} = \frac{V}{I_c}$$

$$= r_c(1-\alpha)+\frac{\alpha r_c r_e}{R_s+r_b+r_e}\,.$$

Similarly, relations may be derived for common-base and common-collector configurations. After certain approximations the results are as given in Table 3.

The T parameters may be related to the h parameters. For the common-emitter circuit,

$$V_{be} = I_b(r_b+r_e)+I_cr_e\,,$$

$$V_{ce} = I_c(r_c+r_e)-\alpha(I_c+I_b)r_c+I_br_e\,.$$

TABLE 3

	Common base	Common emitter	Common collector
Z_{in}	$r_e+r_b(1-\alpha)$	$r_b+\dfrac{r_e}{(1-\alpha)}$	$r_b+\dfrac{R_l}{1-\alpha}$
Z_{out}	$r_c-\dfrac{\alpha r_c r_b}{r_e+r_b+Z_s}$	$r_c(1-\alpha)+\dfrac{\alpha r_c r_e}{r_e+r_b+Z_s}$	$r_e+(R_s+r_b)(1-\alpha)$
Current gain	$-\alpha$	$\dfrac{\alpha}{1-\alpha}$	$-\dfrac{1}{1-\alpha}$
Voltage gain	$\dfrac{\alpha R_l}{r_e+r_b(1-\alpha)}$	$-\dfrac{\alpha R_l}{r_e+r_b(1-\alpha)}$	$\dfrac{R_l}{R_l+r_b(1-\alpha)}$
Power gain	$\dfrac{\alpha^2 R_l}{r_e+r_b(1-\alpha)}$	$\dfrac{\alpha^2 R_l}{(1-\alpha)\,[r_e+r_b(1-\alpha)]}$	$\dfrac{1}{1-\alpha}$

Therefore $$I_c = \frac{(\alpha r_c - r_e)}{r_e + r_c(1-\alpha)} I_b + \frac{V_{ce}}{r_e + r_c(1-\alpha)}.$$

But $$r_e \ll r_c.$$

Therefore $$I_c = \left(\frac{\alpha}{1-\alpha}\right) I_b + \frac{V_{ce}}{r_c(1-\alpha)}.$$

Comparing this equation with $I_c = h_{fe} I_b + h_{oe} V_{ce}$,

$$h_{fe} = \frac{\alpha}{1-\alpha} \quad \text{or} \quad \alpha = \frac{h_{fe}}{1+h_{fe}},$$

$$h_{oe} = \frac{1}{r_c(1-\alpha)} \quad \text{or} \quad r_c = \frac{1+h_{fe}}{h_{oe}}.$$

Also $$V_{be} = I_b(r_b + r_e) + \left(\frac{\alpha}{1-\alpha} I_b + \frac{V_{ce}}{r_c(1-\alpha)}\right) r_e$$

$$= I_b\left(r_b + \frac{r_e}{1-\alpha}\right) + \frac{r_e}{r_c(1-\alpha)} V_{ce}.$$

By comparison with $V_{be} = h_{ie}I_b + h_{re}V_{ce}$,

$$h_{re} = \frac{r_e}{r_c(1-\alpha)} \quad \text{or} \quad r_e = \frac{h_{re}}{h_{oe}}$$

and
$$h_{ie} = r_b + \frac{r_e}{1-\alpha} \quad \text{or} \quad r_b = h_{ie} - \frac{h_{re}(1+h_{fe})}{h_{oe}}.$$

Modifications may be made to the "T" circuit to enable it to be used at high frequencies. These follow the same lines as the modifications made to the "h" equivalent circuit and for the common-base configuration a high-frequency equivalent circuit is as shown in Fig. 5.19.

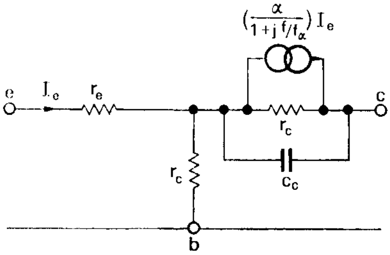


FIG. 5.19. High-frequency "T" equivalent circuit.

5.4. The Hybrid-π Equivalent Circuit

The common-emitter hybrid-π equivalent circuit may be derived from the equivalent "T" circuit by means of the current generator theorem shown in Fig. 5.20. The theorem shows that a current

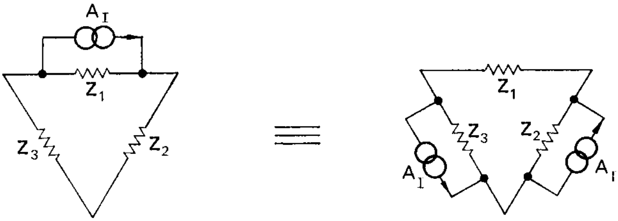


FIG. 5.20. Current generator theorem.

generator across one of the three impedances connected in delta may be replaced by identical current generators across the other two. The theorem may easily be verified by calculating the voltage across any impedance for both configurations.

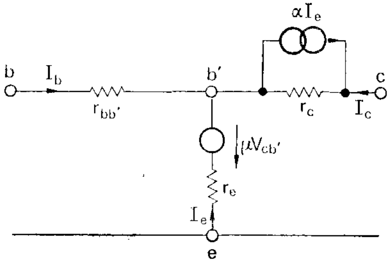


FIG. 5.21. Common-emitter equivalent "T" circuit.

Reorientating Fig. 5.12 gives the common-emitter equivalent circuit shown in Fig. 5.21. From this diagram $I_e = \left(\frac{V_{eb'} - \mu V_{cb'}}{r_e}\right)$, and this current may be substituted in the current generator αI_e. Also $\mu V_{cb'}$ in series with r_e may be replaced by the equivalent current generator to give Fig. 5.22. The current generators $\frac{\mu V_{cb'}}{r_e}$ and $\frac{\alpha V_{cb'}}{r_e}$ may each be replaced by two others to give the equivalent

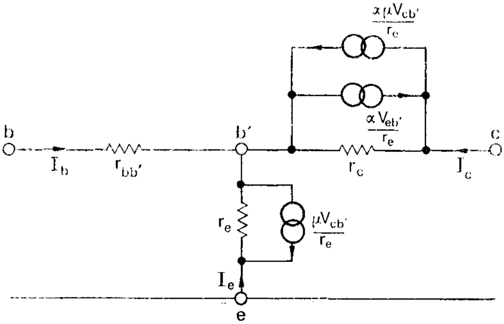


FIG. 5.22. Development of hybrid-π equivalent circuit—stage 1.

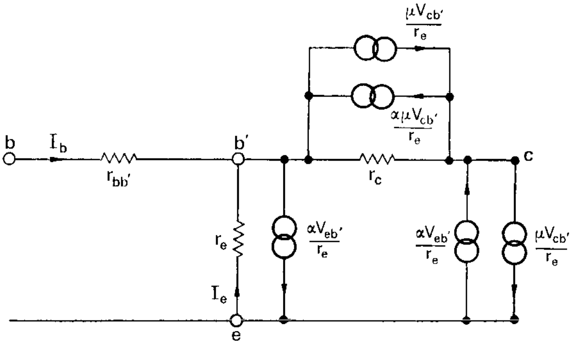


FIG. 5.23. Development of hybrid-π equivalent circuit—stage 2.

circuit shown in Fig. 5.23, and further reduced by converting some of the current generators to resistors with the appropriate voltages across them.

For the link eb', $\frac{\alpha V_{eb'}}{r_e}$ may be replaced by a resistance $-r_e/\alpha$ which in parallel with r_e gives $\frac{r_e(-r_e/\alpha)}{r_e - r_e/\alpha} = \frac{r_e}{1-\alpha}$, denoted by $r_{b'e}$. For cb' the current generators may be combined to give a single generator $\frac{\mu(1-\alpha)V_{cb'}}{r_e}$ flowing from b' to c, and since $\mu = \frac{r_e}{2r_c(1-\alpha)}$, this may be replaced by a current generator $\frac{V_{cb'}}{2r_c}$ and hence by a resistor $-2r_c$. This resistor in parallel with r_c gives a resistance of $2r_c$ denoted by $r_{b'c}$. In practice, r_e is low and $V_{cb'}$ is approximately equal to V_{ce} so that the current generator $\mu V_{cb'}/r_e$ may be replaced by a resistance $r_e/\mu = 2r_e(1-\alpha) = r_{ce}$. Using these substitutions the hybrid equivalent circuit may be derived as shown in Fig. 5.24, where $g_m = \alpha/r_e$.

This form of the grounded-emitter equivalent circuit is particularly useful at high frequencies when the effective base–emitter capacitance $C_{b'e}$ may be shown across $r_{b'e}$ and collector–base capacitance $C_{b'c}$ across $r_{b'c}$ as illustrated by Fig. 5.25.

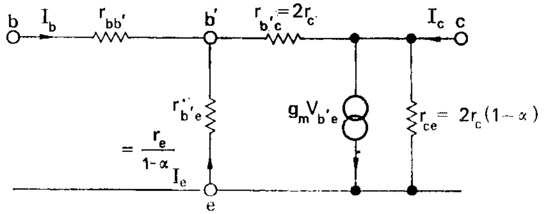


FIG. 5.24. Hybrid-π equivalent circuit.

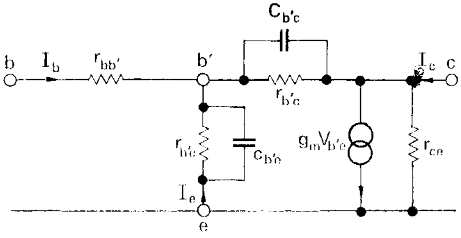


FIG. 5.25. High-frequency hybrid-π equivalent circuit.

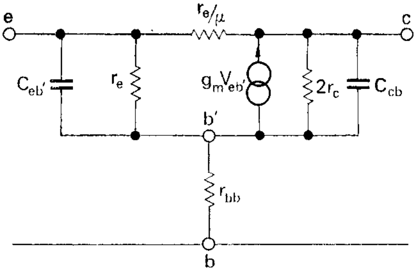


FIG. 5.26. Common-base hybrid-π equivalent circuit.

A similar argument may be used with the common-base configuration and gives rise to the equivalent circuit shown in Fig. 5.26.

Analysis using the hybrid-π circuit may be illustrated by considering an amplifier with a load impedance Z_l (or admittance Y_l). Ignoring bias components the equivalent circuit of such a stage is shown in Fig. 5.27. In this circuit $Z_{b'c}$ includes $r_{b'c}$ and $C_{b'c}$; $Z_{b'e}$ includes $r_{b'e}$ and $C_{b'e}$ and $g_{ce} = 1/r_{ce}$.

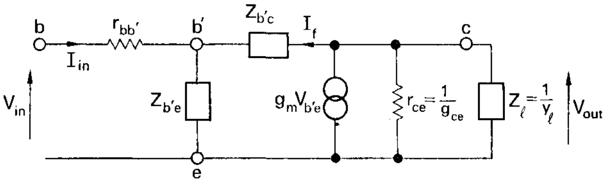


FIG. 5.27. Analysis using hybrid-π equivalent circuit.

To a first approximation, $V_{\text{out}} = \left(\dfrac{g_m V_{b'e}}{g_{ce}+Y_l}\right)$ is much greater than $V_{b'e}$, and the feedback current is given by

$$I_f = -\frac{g_m V_{b'e}}{(g_{ce}+Y_l)Z_{b'c}}.$$

Therefore
$$V_{b'e} = \left(I_{\text{in}} - \frac{g_m V_{b'e}}{(g_{ce}+Y_l)Z_{b'c}}\right) Z_{b'e}$$

or
$$V_{b'e} = \frac{I_{\text{in}} Z_{b'e}}{1+\dfrac{g_m Z_{b'e}}{(g_{ce}+Y_l)Z_{b'c}}}.$$

Therefore
$$\begin{aligned} V_{\text{in}} &= I_{\text{in}} r_{bb'} + V_{b'e} \\ &= I_{\text{in}} r_{bb'} + \frac{I_{\text{in}} Z_{b'e}}{1+\dfrac{g_m Z_{b'e}}{(g_{ce}+Y_l)Z_{b'c}}}. \end{aligned}$$

Hence input impedance
$$\begin{aligned} Z_{\text{in}} &= \frac{V_{\text{in}}}{I_{\text{in}}} \\ &= r_{bb'} + \frac{Z_{b'e}}{1+\dfrac{g_m Z_{b'e}}{(g_{ce}+Y_l)Z_{b'c}}}. \end{aligned}$$

The current gain of the circuit is given by

$$\begin{aligned} A_I &= -\frac{g_m\left(\dfrac{Y_l}{g_{ce}+Y_l}\right)V_{b'e}}{\left(1+\dfrac{g_m Z_{b'e}}{(g_{ce}+Y_l)Z_{b'c}}\right)\dfrac{V_{b'e}}{Z_{b'e}}} \\ &= -\frac{g_m Y_l Z_{b'e} Z_{b'c}}{(g_{ce}+Y_l)Z_{b'c}+g_m Z_{b'e}}. \end{aligned}$$

Then voltage gain $$A = \frac{I_{\text{out}} Z_l}{I_{\text{in}} Z_{\text{in}}}$$

$$= -\frac{g_m Z_{b'e} Z_{b'c}}{(g_{ce}+Y_l)Z_{b'c}+g_m Z_{b'e}}$$

$$\times \frac{1}{r_{bb'}+\left\{Z_{b'e}\Big/\left[1+\frac{g_m Z_{b'e}}{(g_{ce}+Y_l)Z_{b'c}}\right]\right\}}$$

$$= -\frac{g_m Z_{b'e} Z_{b'c}}{(g_{ce}+Y_l)\, Z_{b'c}(r_{bb'}+Z_{b'e})+g_m Z_{b'e} r_{bb'}}\,.$$

To determine the output admittance, a voltage V may be applied to the output circuit as shown in Fig. 5.28. The current fed back equals $V/Z_{b'c}$, very nearly.

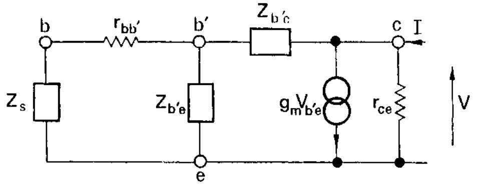


FIG. 5.28. Output admittance using hybrid-π equivalent circuit.

Hence, $$V_{b'e} = \frac{V}{Z_{b'c}} \times \frac{Z_{b'e}(r_{bb'}+Z_s)}{Z_{b'e}+r_{bb'}+Z_s}\,.$$

Also current from source

$$I = V g_{ce} + \frac{V}{Z_{b'c}} + \frac{g_m V Z_{b'e}(r_{bb'}+Z_s)}{Z_{b'c}(Z_{b'e}+r_{bb'}+Z_s)}\,.$$

Hence the output admittance is given by

$$Y_{\text{out}} = g_{ce} + \frac{1}{Z_{b'c}} + \frac{g_m Z_{b'e}(r_{bb'}+Z_s)}{Z_{b'c}(Z_{b'e}+r_{bb'}+Z_s)}\,.$$

The power gain of the circuit is given by

$$A_p = (\text{current gain})^2 \times \left(\frac{R_l}{R_{\text{in}}}\right)$$
$$= (\text{voltage gain})^2 \times \left(\frac{G_l}{G_{\text{in}}}\right).$$

Although both hybrid-π and "y"-equivalent circuits may be used at high frequencies, there appears to be definite regions where one is more suitable than the other. Manipulation of the circuit equations is easier with the y parameters, but difficulties arise when variable frequency or wide-band circuits are considered. The algebra of the hybrid-π equivalent circuit is usually more complex but, since the parameters are independent of frequency, the circuit may be used over a wide frequency range.

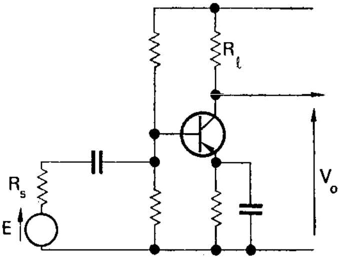


FIG. 5.29. Wide-band amplifier.

A particular example of the use of hybrid-π equivalent circuit is the wide-band amplifier shown in Fig. 5.29. Assuming a resistive source impedance and ignoring the effect of bias components, the equivalent circuit of the amplifier may be represented by Fig. 5.30. Also in practice, since wide-band amplifiers incorporate relatively low collector loads and have limited gain, r_{ce} may be ignored by comparison with R_l and the feedback current may be neglected. The equivalent circuit may then be reduced to that shown in Fig. 5.31.

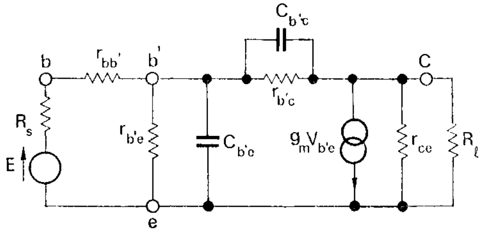


FIG. 5.30. Equivalent circuit of Fig. 5.29.

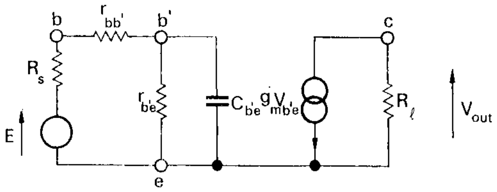


FIG. 5.31. Simplified equivalent circuit of Fig. 5.29.

The overall voltage gain of the circuit is then obtained:

$$V_{\text{out}} = -\frac{g_m V_{b'e} R_l}{E}$$

$$= -\frac{g_m R_l}{E} \frac{E\left(\dfrac{r_{b'e}/j\omega C_{b'e}}{r_{b'e}+1/j\omega C_{b'e}}\right)}{R_s + r_{bb'} + \dfrac{r_{b'e}/j\omega C_{b'e}}{r_{b'e}+(1/j\omega C_{b'e})}}.$$

Therefore $$A_V = \frac{V_{\text{out}}}{E} = -\frac{g_m R_l r_{b'e}/j\omega C_{b'e}}{(R_s + r_{bb'})\left(r_{b'e} + \dfrac{1}{j\omega C_{b'e}}\right) + \dfrac{r_{b'e}}{j\omega C_{b'e}}}$$

$$= -\frac{g_m R_l r_{b'e}}{(R_s + r_{bb'} + r_{b'e}) + j\omega C_{b'e} r_{b'e}(R_s + r_{bb'})}.$$

At the 3 dB point the real and quadrature terms in the denominator are equal,

i.e. $$f_{3\,\text{dB}} = \frac{(R_s + r_{bb'} + r_{b'e})}{2\pi C_{b'e} r_{b'e}(R_s + r_{bb'})}.$$

The source resistance obviously has a considerable effect on the bandwidth. If a constant current source is used, R_s is very high and the 3 dB point is given by

$$f_{3\,\mathrm{dB}} = \frac{1}{2\pi C_{b'e} r_{b'e}}.$$

If a voltage source is used, R_s is very low and

$$f_{3\,\mathrm{dB}} = \frac{(r_{bb'} + r_{b'e})}{2\pi C_{b'e} r_{b'e} r_{bb'}}.$$

Normally such stages are operated at low collector currents when $r_{bb'}$ is considerably less than $r_{b'e}$. The 3 dB cut-off point is then given by

$$f_{3\,\mathrm{dB}} = \frac{1}{2\pi C_{b'e} r_{bb'}}.$$

This frequency is usually much higher than when a constant current source is used.

The above expressions also depend on the values of $r_{b'e}$ and $C_{b'e}$. Although these are independent of frequency, they do, however, depend on the standing emitter current, i.e. $r_{b'e}$ is inversely proportional to the emitter current and $C_{b'e}$ is proportional to the emitter current.

Taking some typical values;

$$R_s = 50\,\Omega, \quad r_{bb'} = 50\,\Omega, \quad r_{b'e} = 1\ \mathrm{k}\Omega, \quad R_l = 1\ \mathrm{k}\Omega,$$
$$r_{ce} = 20\ \mathrm{k}\Omega,$$
$$g_m = 30\ \mathrm{mA/V}, \quad r_{b'c} = 1\ \mathrm{M}\Omega.$$

At low frequencies, ignoring the feedback resistance,

$$I_{\mathrm{in}} = \frac{E}{1100} = 0{\cdot}91\ \mathrm{mA},$$
$$V_{b'e} = 0{\cdot}91E,$$
$$V_{\mathrm{out}} = -30 \times \tfrac{20}{21} \times 0{\cdot}91E = -26E,$$

i.e. voltage gain, $A_V = -26$ times.

$$\text{The feedback current} = \frac{E-(-26\,E)}{10^6}$$

$$= 0{\cdot}027 \text{ mA.}$$

Thus the error in ignoring the feedback current is only about 3%, which is probably less than the variation in parameter values in a particular group of transistors of the same type. A typical value of $C_{b'c}$ is 5000 pF, giving a bandwidth of

$$f = \frac{1100}{2\pi \times 5000 \times 10^{-12} \times 10^3 \times 10^2}$$

$$= 350 \text{ kHz.}$$

At this frequency $C_{b'c}$ would be about 10 pF, i.e. its reactance is about 45·5 kΩ. The current fed back is approximately $\frac{26/\sqrt{(2)}}{45{\cdot}5}E = 0{\cdot}4E$ mA compared with an input current of $\frac{E}{\left|100+\frac{1000\times j91}{1000+j91}\right|} \simeq 7{\cdot}1E$ mA. The error is only about 5%. Another use of high-frequency equivalent circuits is to determine unilateralizing components. For the circuit shown in Fig. 5.32 an equivalent circuit as shown in Fig. 5.33 may be drawn. Using the values indicated $y_{bb'} = 0{\cdot}01$ S, $y_{b'e} = 0{\cdot}001 + j\,0{\cdot}00314$ S and $y_{b'c} = 0{\cdot}33 \times 10^{-6} + j\,31{\cdot}4 \times 10^{-6}$ S. Then, converting to a normal π as illustrated by Fig. 5.34,

$$y_{bc} = \frac{y_{bb'}y_{b'c}}{y_{bb'}+y_{b'e}+y_{b'c}} = 7{\cdot}87 \times 10^{-6} + j26{\cdot}4 \times 10^{-6} \text{ S.}$$

At 500 kHz this is equivalent to 10,400+j 3500 Ω, i.e. a resistance of 10·4 kΩ in series with 9·1 pF. If the closely coupled transformer has a turns ratio of $\sqrt{(5)} : 1$, the 180° phase shift transformer enables unilateralizing components C_n and R_n of 45·5 pF and

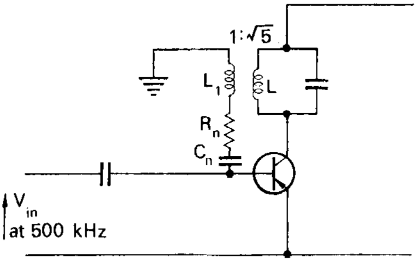


FIG. 5.32. Unilateralized amplifier.

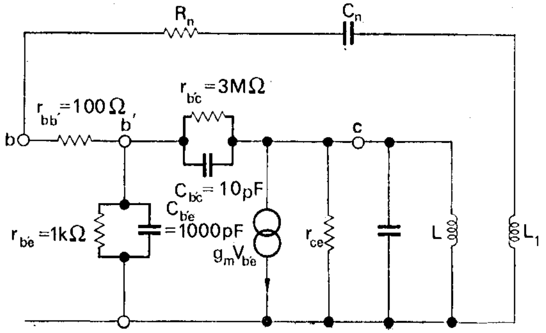


FIG. 5.33. Equivalent circuit of Fig. 5.32.

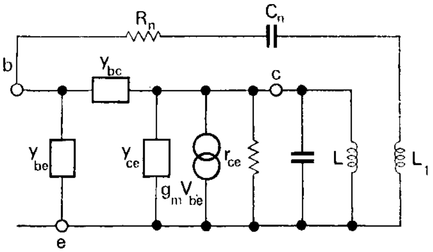


FIG. 5.34. Modified equivalent circuit of Fig. 5.32.

2·08 kΩ to be incorporated. Under these conditions the energy fed back externally exactly balances that fed back internally through the transistor.

Questions on Chapter 5

Q.1. The hybrid parameters for a transistor used in the common-emitter configuration are:

$$h_{ie} = 1{\cdot}5\,\text{k}\Omega, \quad h_{re} = 10^{-4}, \quad h_{fe} = 70, \quad h_{ve} = 100\,\mu\text{S}.$$

The transistor is fed from a source of 1 kΩ resistance and the load resistor is also 1 kΩ.

Calculate (a) the input resistance, (b) the output resistance, (c) the voltage gain, and (d) the current gain.

Q.2. Show that the image impedances of a transistor represented by the "h" equivalent circuit are given by:

$$Z_1 = \sqrt{\left(\frac{h_i}{h_o h}\right)} \quad \text{and} \quad Z_2 = \sqrt{\left(\frac{hh_i}{h_o}\right)},$$

where $$h = h_o h_i - h_r h_f.$$

Q.3. Derive expressions for the common-emitter y parameters in terms of the common-base y parameters.

Q.4. Explain what is meant by the y parameters of a transistor. The collector load of a unilateralized transistor, operating at 10 MHz, is a tuned circuit tightly coupled (i.e. unity coefficient of coupling) to a pure resistance of 500 Ω. The y parameters of the transistor are

$$y_{ie} = (20+j\,15)\text{ mS}, \qquad y_{re} = 0{\cdot}4 \angle -100\text{ mS},$$
$$y_{fe} = 80 \angle -70\text{ mS}, \qquad y_{oe} = (2+j\,6)\text{ mS}.$$

Estimate the turns ratio of the transformer for maximum power transfer and the value of the unilateralizing components.

Q.5. Derive expressions for the input and output impedance of common-base and common-collector circuits in terms of the T parameters and the source and load resistances.

Q.6. The parameters of a transistor used as a grounded emitter a.f. amplifier are:

$$r_e = 50\,\Omega, \quad r_b = 500\,\Omega, \quad r_c = 1\text{ M}\Omega, \quad \alpha = 0{\cdot}98.$$

Determine from first principles the input and output resistances of the stage assuming a load resistance of 8 kΩ and a generator resistance of 850 Ω.

Q.7. Explain quantitatively how the bandwidth of transistor in the common-emitter configuration depends on the source impedance. The hybrid-π equivalent circuit may be assumed.

Q.8. Show that if internal feedback is ignored the optimum value of the inductance to be inserted in series with the load resistance R_L for maximum bandwidth is given by

$$L = \frac{R_L C_{b'e} r_{b'e}(r_s + r_{bb'})}{R_s + r_{bb'} + r_{b'e}}.$$

CHAPTER 6

Charge Control of Transistors

A BASIC difference between bipolar transistors and vacuum valves or field effect transistors is the fact that they are current, instead of voltage, controlled. It is a small step to think of bipolar transistors as charge-controlled devices. This approach is useful when considering the switching performance of transistors.

6.1. Charge Distribution in Bipolar Transistors

To illustrate how the charge in a transistor, particularly in the base region, varies, this section will be concerned with a device, in the common-emitter connection, taken from an initial cut-off condition into saturation.

6.1.1. The cut-off condition

In this condition, both junctions of the transistor are reverse-biased, the only current flowing is the leakage current, and the base charge is negligible.

6.1.2. Transistor at the onset of current

Before current flows in a transistor the emitter junction must be forward-biased and base current must be supplied to charge the emitter–base depletion-layer capacitance and, incidentally, the collector–base depletion-layer capacitance. The charges required are denoted by Q_E and Q_C respectively, where Q_E is a function of the initial reverse bias on the emitter junction and Q_C a function of the initial bias on the collector junction. The charge distribution in the base is then as shown in Fig. 6.1, and is due to majority carriers (i.e. electrons for an *n*-type base region).

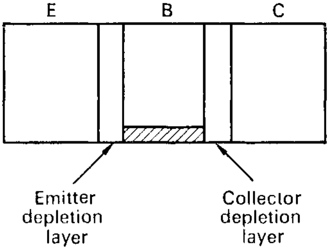


FIG. 6.1. Charge on depletion layers.

6.1.3. *Forward-biased transistor*

In this condition an emitter current flows and minority carriers diffuse across the base region (i.e. holes in a *pnp* and electrons in an *npn* transistor). This, of course, must be balanced by an influx of majority carriers to preserve charge neutrality.

The current I flowing through the base follows the usual relation for a diffusion current,

i.e.
$$I = De\frac{dN}{dx}A,$$

where D is the diffusion coefficient, e is the charge on an electron, dN/dx is the charge concentration gradient, and A is the area.

This current, due to minority carriers, indicates a minority charge distribution with a constant charge concentration gradient as indicated in Fig. 6.2. This statement has certain limitations, but is approximately true for abrupt junctions. At the collector junction the carriers are swept across the collector depletion layer. The collector current flowing through the load resistance reduces the collector base voltage and results in a narrower depletion layer. Additional base current will be required to compensate for the change in capacitance and charge stored.

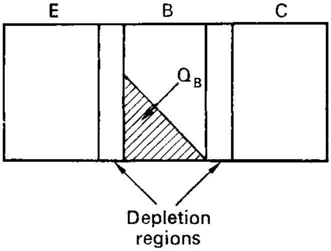


FIG. 6.2. Charge distribution in a forward-biased transistor.

The collector current increases linearly with base charge until the onset of saturation when the collector–base voltage is approximately zero (in fact a slight reverse bias exists). The base charge in this limiting condition is denoted by Q_B.

6.1.4. *Overdriven transistor*

If additional charge is injected into the base region such that it is taken into saturation, there can be no increase in collector current, and excess charge Q_{BS} builds up in the base region. This charge builds up uniformly across the base as shown in Fig. 6.3. Under this condition both collector and emitter regions are forward-biased and both inject carriers into the base.

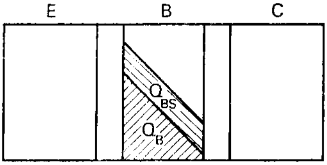


FIG. 6.3. Charge distribution in a saturated transistor.

6.2. Charge Control Parameters

6.2.1. Charge on depletion layers

As stated previously, before a transistor can be switched on from an initial reverse-biased condition, a charge Q must be supplied to both emitter and collector depletion-layer capacitances. Consider a depletion-layer capacitance that varies with bias voltage as shown in Fig. 6.4. If the initial bias is V_1 falling to a final value V_2, the charge supplied is given by

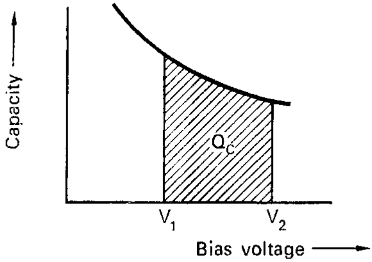


FIG. 6.4. Variation of capacitance of depletion layers.

$$Q = \int_{V_2}^{V_1} C\,dV.$$

In the case of an abrupt junction, $C = KV^{-1/2}$.

Therefore
$$\begin{aligned} Q &= \int_{V_2}^{V_1} KV^{-1/2}\,dV \\ &= 2K[V^{1/2}]_{V_2}^{V_1} \\ &= 2K(\sqrt{V_1}-\sqrt{V_2}). \end{aligned}$$

Normally the collector depletion-layer capacitance is the more important, i.e. $Q_C = 2K(\sqrt{V_1}-\sqrt{V_2})$, and this is often stated by the manufacturer.

To be strictly correct, V_1 and V_2 should include the voltage normally associacted with an unbiased *pn* junction.

Alternatively to a first approximation, the charge on the collector depletion-layer may be written

$$Q_C = MC_C(V_1 - V_2),$$

where C_C is the small-signal depletion-layer capacitance measured at the off collector voltage, and M is a factor between 1 and 2 depending on the type of junction, alloy, or drift, and the magnitude of the voltage change $(V_1 - V_2)$.

6.2.2. Base charge in an unsaturated transistor

When the transistor is operating under active conditions, i.e. in the time between current starting to flow and saturation, to a close approximation the ratio of base charge to the collector current is a constant. This obviously must have the dimensions of time and is measured and stated for the limiting condition just before the onset of saturation. This corresponds to a base charge Q_B and a maximum collector current $I_C (= h_{FE} I_B$, if leakage current is ignored).

Then the collector or fundamental time constant T_C is given by

$$T_C = Q_B / I_C.$$

The collector time constant may be related to the physical constants of the transistor. When the condition illustrated by Fig. 6.5

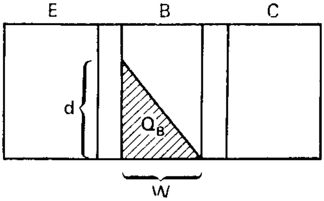


FIG. 6.5. Evaluation of base charge.

is considered, the maximum minority carrier density at the edge of the emitter–base depletion layer is d and falls to zero over a distance equal to the base width W,

i.e. $$\frac{dN}{dx} = \frac{d}{w}.$$

d in this context represents the hole density in a *pnp* and the electron density in an *npn* transistor.

Hence
$$Q_B = \tfrac{1}{2}wAed$$

and $$I_C = De\frac{d}{w}A.$$

Therefore $$T_C = \frac{Q_B}{I_C}$$
$$= \frac{\frac{1}{2}wAed}{De(d/w)A}$$
$$= \frac{w^2}{2D}.$$

Ideally, once the minority carrier concentration gradient has been set up and charge neutrality maintained by the influx of majority carriers, the collector current should flow indefinitely. In practice, however, recombination takes place in the base region, and a base current must flow to compensate for this. The base current flowing is proportional to the stored charge, and the ratio—again usually stated at the edge of saturation—is given by

$$T_B = \frac{Q_B}{I_B},$$

where T_B is the base time constant—i.e. the minority carrier lifetime in the base and $I_B = I_C/h_{FE}$, where h_{FE} is the large-signal current gain.

Hence, ignoring leakage current,

$$\begin{aligned} T_B &= \frac{Q_B}{I_B} = \frac{h_{FE}Q_B}{I_C} \\ &= h_{FE}T_C \\ &= \frac{h_{FE}W^2}{2D}. \end{aligned}$$

6.2.3. *Switch-on charge*

The switch-on charge (Q_{on}) may be defined as the base charge required to bring the transistor from the initial off condition to the verge of saturation,

i.e. $$Q_{\text{on}} = Q_B + Q_C + Q_E.$$

If, as is normally valid for an abrupt junction transistor, Q_E may be ignored,

$$Q_{\text{on}} = I_C T_C + Q_C.$$

6.2.4. *Charge in a saturated transistor*

Once a transistor is saturated and the collector current is at its maximum value, any increase in base current results in an increase Q_{BS} in the stored base charge. This is limited in practice by recombination in the base region, the excess base current I_{BS} compensating for the minority charge carriers required to maintain the excess charge Q_{BS} at a constant value. The ratio of Q_{BS} to I_{BS} is the saturation time constant T_S giving the effective lifetime of the excess charge carriers in the base region.

Hence

$$T_S = \frac{Q_{BS}}{I_{BS}},$$

where $$I_{BS} = I_B - \frac{I_C}{h_{FE}}.$$

The total charge in the base is then given by

$$\begin{aligned} Q_{\text{total}} &= Q_{\text{on}}+Q_{BS} \\ &= I_C T_C+Q_C+I_{BS}T_S. \end{aligned}$$

6.2.5. Switch-off charge

When an overdriven transistor has to be switched off, the total charge that must be extracted from the base is given by

$$\begin{aligned} Q_{\text{off}} &= Q_{\text{on}}+Q_{BS} \\ &= I_C T_C+Q_C+I_{BS}T_S \\ &= I_C T_C+Q_C+\left(I_B-\frac{I_C}{h_{FE}}\right)T_S. \end{aligned}$$

To give an idea of the magnitudes of the charges involved, consider a transistor operating from a 6 V supply through a 600 Ω load resistance. The relevant time constant are typically $T_C = 30$ ns and $T_S = 800$ ns, h_{FE} is assumed to be 25, the base current 0·75 mA, and the collector capacitance 10 pF at 6 V. Q_E is assumed to be negligible.

$$C_C = KV^{-1/2}.$$

Therefore $$K = C_C V^{1/2} = 10^{-11}\sqrt{6}.$$

Assuming the collector base voltage falls to zero,

$$\begin{aligned} Q_C &= 2K[\sqrt{(6)}-0] \\ &= 12\times10^{-11} \\ &= 120 \text{ pC}. \end{aligned}$$

With the same assumption of $V_{CB} = 0$,

$$I_C = \frac{6}{600}\text{ A} = 10\text{ mA}.$$

Therefore $$Q_B = I_C T_C$$
$$= 10 \times 10^{-3} \times 30 \times 10^{-9} \text{ C}$$
$$= 300 \text{ pC}.$$

Therefore $$Q_{\text{on}} = 300 + 120$$
$$= 420 \text{ pC}.$$

The excess base current $$I_{BS} = I_B - \frac{I_C}{h_{FE}}$$
$$= 0{\cdot}75 - \frac{10}{25}$$
$$= 0{\cdot}35 \text{ mA}.$$

Therefore $$Q_{BS} = 0{\cdot}35 \times 10^{-3} \times 800 \times 10^{-9} \text{ C}$$
$$= 280 \text{ pC}.$$

Therefore $$Q_{\text{off}} = 420 + 280$$
$$= 700 \text{ pC}.$$

6.3. Transient Parameters

To illustrate the transient parameters involved when a transistor is switched from the off to the on position, consider a common-emitter *npn* transistor, initially cut off, receiving a pulse of current on the base sufficient to saturate the device. The configuration is shown in Fig. 6.6, and typical input and output current waveforms are illustrated in Fig. 6.7.

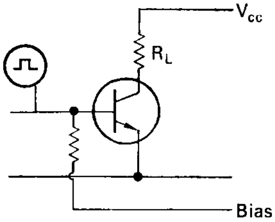


FIG. 6.6. Pulse amplifier.

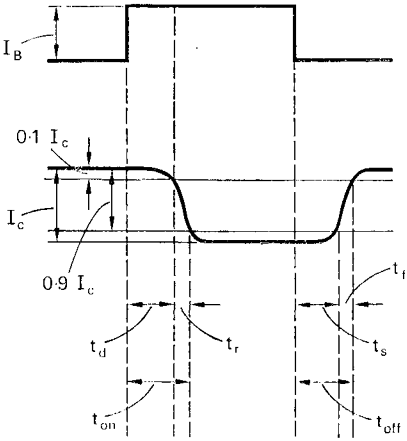


FIG. 6.7. Waveform for pulse amplifier.

After receiving the pulse no output will be obtained until the depletion regions have been charged. This corresponds to supplying a charge Q_E+Q_C by means of a current I_B,

i.e. $$\text{delay time } t_d = \frac{Q_E+Q_C}{I_B}.$$

In practical measurements there are no sharp corners in the output pulse, and t_d is measured from the beginning of the input pulse to a decrease of 10% of the final (output pulse) amplitude.

As the output pulse rises from zero towards a value of V_{CC}/R_L, the minority carrier density increases at a rate dq/dr as indicated in Fig. 6.8. This must be balanced by majority carriers flowing

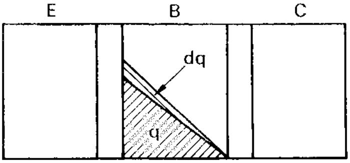


FIG. 6.8. Incremental base charge.

into the base. In addition, majority carriers must flow to compensate the recombination in the base region and to allow for the change in charge on the collector depletion layer.

To a first approximation the latter may be ignored, giving at a particular value of base charge q

$$I_B = \frac{q}{T_B} + \frac{dq}{dt}. \tag{6.3.1}$$

As before, due to the output pulse being rounded off, specific reference points must be chosen. A practical value for the rise time is the time taken for the output pulse to rise from 10 to 90% of its final amplitude. This may be estimated by solving the differential equation (6.3.1). Taking Laplace transforms of eqn. (6.3.1),

$$\frac{I_B}{s} = \frac{\bar{q}}{T_B} + s\bar{q} - q_o.$$

But $q_o = 0$ when the output pulse starts to rise. Therefore

$$\begin{aligned}\bar{q} &= \frac{I_B}{s} \times \frac{1}{[s+(1/T_B)]}\\ &= I_B\left(\frac{T_B}{s} - \frac{T_B}{s+(1/T_B)}\right)\\ &= I_B T_B\left(\frac{1}{s} - \frac{1}{s+(1/T_B)}\right).\end{aligned}$$

Taking the inverse,

$$q = I_B T_B(1 - \varepsilon^{-t/T_B}).$$

Hence the instantaneous collector current is given by

$$\begin{aligned}i_C &= \frac{q}{T_C}\\ &= \frac{h_{FE}q}{T_B}\\ &= h_{FE} I_B(1 - \varepsilon^{-t/T_B}).\end{aligned}$$

This expression indicates an exponential rise of collector current that reaches a value i_C in a time given by

$$\varepsilon^{-t/T_B} = 1 - \frac{i_C}{h_{FE}T_B}$$

or

$$t = T_B \log_\varepsilon \frac{1}{[1-(i_C/h_{FE}I_B)]}.$$

But i_C rises to a maximum value of V_{CC}/R_L, and hence the time taken to rise from $0{\cdot}1V_{CC}/R_L$ to $0{\cdot}9V_{CC}/R_L$, i.e. the rise time is given by

$$t_r = T_B \log_\varepsilon \frac{1}{\left(1-\frac{0{\cdot}9V_{CC}}{h_{FE}I_BR_L}\right)} - T_B \log_\varepsilon \frac{1}{\left(1-\frac{0{\cdot}1V_{CC}}{h_{FE}I_BR_L}\right)}$$

or

$$t_r = T_B \log_\varepsilon \left(\frac{1-\left(\frac{0{\cdot}1\mathrm{V}_{CC}}{h_{FE}I_BR_L}\right)}{1-\left(\frac{0{\cdot}9V_{CC}}{h_{FE}I_BR_L}\right)}\right).$$

From the delay time and the rise time the total time required to switch the transistor on t_{on} may be estimated,

i.e.

$$t_{\text{on}} = t_d + t_r.$$

While the base current is maintained at the value I_B (greater than I_C/h_{FE}), the base charge builds up until limited by recombination, but the output current remains at $I_C = V_{CC}/R_L$.

After the input pulse falls to zero, there is a delay before the output current falls to zero. This delay is due to the time taken for the excess charge to recombine and is termed the storage time, denoted by t_s. As before, some definite reference point must be chosen for practical measurements, and the storage time is measured from the time the input pulse falls to zero and the time the output pulse falls to 90% of its maximum value.

During the storage time the base current is zero while the collector current is initially constant at $I_C(= V_{CC}/R_L)$ but finally falls to $0{\cdot}9V_{CC}/R$. Consider, firstly the period when I_C is constant and the excess base charge decreases to zero. The excess base charge decays at a rate (dq_{BS}/dt) due to recombination of excess charge carriers (q_{BS}/T_s) and to recombination of the base charge required for saturation (Q_B/T_B)

i.e.
$$0 = \frac{Q_B}{T_B} + \frac{q_{BS}}{T_s} + \frac{dq_{BS}}{dr}.$$

But
$$Q_B = I_C T_C = \frac{V_{CC}}{R_L} T = \frac{V_{CC}}{R_L} \cdot \frac{T_B}{h_{FE}}.$$

Therefore
$$0 = \frac{V_{CC}}{R_L h_{FE}} + \frac{q_{BS}}{T_s} + \frac{dq_{BS}}{dt}.$$

Taking Laplace transforms,

$$0 = \left(\frac{V_{CC}}{R_L h_{FE}}\right)\Big/ s + \frac{\bar{q}_{BS}}{T_s} + s\bar{q}_{BS} - q_{BSo}.$$

But at $t = 0$, $q_{BSo} = I_{BS}T_s = \left(I_B - \frac{V_{CC}}{R_L h_{FE}}\right) T_s.$

Therefore
$$\bar{q}_{BS}\left(s + \frac{1}{T_s}\right) = \left(I_B - \frac{V_{CC}}{R_L h_{FE}}\right) T_s - \frac{V_{CC}}{R_L h_{FE} s}$$
$$= I_B T_s - \frac{V_{CC}}{R_L h_{FE}} \left\{T_s + \frac{1}{s}\right\}$$

or
$$\bar{q}_{BS} = \frac{I_B T_s}{(s + (1/T_s)]} - \frac{V_{CC}}{R_L h_{FE}} \frac{T_s}{s}.$$

Taking the inverse,

$$q_{BS} = I_B T_s \varepsilon^{-t/T_s} - \frac{V_{CC} T_s}{R_L h_{FE}},$$

when $q_{BS} = 0$, $$\varepsilon^{-t/T_s} = \frac{V_{CC}}{R_L h_{FE} I_B}$$

or $$t = T_s \log_\varepsilon \frac{R_L h_{FE} I_B}{V_{CC}}.$$

Once the excess base charge has recombined the base charge, Q_B starts to fall.

Then

$$0 = \frac{q_B}{T_B} + \frac{dq_B}{dt}.$$

Taking Laplace transforms,

$$0 = \frac{\bar{q}_B}{T_B} + s\bar{q}_B - q_{Bo}.$$

But q_{Bo} (the charge at $t = 0$) $= Q_B = \frac{V_{CC}}{R_L h_{FE}} T_B$.

Therefore $$\bar{q}_B = \left(\frac{V_{CC} T_B}{R_L h_{FE}}\right) \Big/ [s + (1/T_B)].$$

Taking the inverse,

$$q_B = \frac{V_{CC} T_B}{R_L h_{FE}} \varepsilon^{-t/T_B}$$

and $$i_C = \frac{q_B}{T_C} = \frac{V_{CC}}{R_L} \varepsilon^{-t/T_B}.$$

When i_C falls to $\frac{0{\cdot}9 V_{CC}}{R_L}$,

$$t = T_B \log_\varepsilon \frac{1}{0{\cdot}9}$$

$$= 0{\cdot}105\, T_B.$$

The storage time t_s, i.e. the time from the end of the input pulse to the time when the collector current falls to 90% of its maximum value, is then given by

$$t_s = 0{\cdot}105 T_B + T_s \log_\varepsilon \frac{R_L h_{FE} I_B}{V_{CC}}.$$

The collector current will continue falling exponentially after the end of the measured storage time obeying the equation derived previously,

i.e. $$i_C = \frac{V_{CC}}{R_L}\varepsilon^{-t/T_B}.$$

The fall time t_f, defined as the time for the output pulse to fall from 90% to 10% of its maximum amplitude is then given by

$$\begin{aligned} t_f &= T_B \log_\varepsilon \frac{1}{0{\cdot}1} - T_B \log_\varepsilon \frac{1}{0{\cdot}9} \\ &= 2{\cdot}3 T_B - 0{\cdot}105 T_B \\ &= 2{\cdot}2 T_B, \end{aligned}$$

This value does not include the change of charge on the collector depletion layer, but the error is small.

Finally, the total turn-off time may be estimated from the relation

$$\begin{aligned} t_{\text{off}} &= t_s + t_f \\ &= 2{\cdot}3 T_B + T_s \log_\varepsilon \frac{R_L h_{FE} I_B}{V_{CC}}. \end{aligned}$$

6.4. Effect of Base Current on Transient Parameters

As shown in the previous section, if a common-emitter transistor is switched on by a current pulse, the rise time t_r is given by

$$t_r = T_B \log_\varepsilon \left(\frac{1 - \dfrac{0{\cdot}1 V_{CC}}{h_{FE} I_B R_L}}{1 - \dfrac{0{\cdot}9 V_{CC}}{h_{FE} I_B R_L}} \right).$$

The greater the base current, i.e. the more the transistor is driven into saturation, the lower will be the rise time. The limiting condition is when the transistor is just saturated,

i.e. $$I_C = h_{FE}I_B \quad \text{(ignoring leakage current)}$$

or $$\frac{V_{CC}}{R_L} = h_{FE}I_B.$$

Therefore $$\frac{V_{CC}}{h_{FE}R_LI_B} = 1.$$

Under this condition, $$t_r = T_B \log_\varepsilon\left\{\frac{0{\cdot}9}{0{\cdot}1}\right\} = 2{\cdot}2T_B.$$

Also the storage time was given by

$$t_s = 0{\cdot}105T_B + T_s \log_\varepsilon \frac{R_Lh_{FE}I_B}{V_{CC}}.$$

The latter expression shows that the greater I_B the greater is the storage time. Hence although faster rise times may be obtained by overdriving the transistor, this is at the expense of the storage time. In the limiting condition when the transistor is just saturated,

$$I_C = h_{FE}I_B = \frac{V_{CC}}{R_L},$$

i.e. $$\frac{R_Lh_{FE}I_B}{V_{CC}} = 1.$$

Then the storage time is at its minimum value of $0{\cdot}105T_B$.

Questions on Chapter 6

Q.1. A transistor in the common-emitter connection has the following parameters:

$$T_C = 0{\cdot}05\ \mu\text{S}, \qquad T_s = 1\ \mu\text{S},$$

$$Q_C = 200\ \text{pC}, \qquad I_C = 10\ \text{mA},$$

$$I_B = 1{\cdot}2\ \text{mA}, \qquad h_{FE} = 50.$$

Estimate the charge that has to be extracted from the base to turn the device off.

Q.2. The d.c. supply to the base of a saturated common-emitter transistor is suddenly reduced from I_{B_1} to I_{B_2}, where I_{B_2} is less than the base current required to saturate the transistor. Show that the time elapsing between the input current being reduced and the collector current falling is given by

$$t = T_B \log_\varepsilon \left[\frac{I_{B_1} - I_{B_2}}{I_B - I_{B_2}}\right],$$

where I_B is the current required just to produce saturation.

Q.3. A saturated common-emitter *pnp* transistor delivers a collector current I_C. The base current is suddenly changed from a value I_B out of the base to a current I_{B_1} into the base. Estimate the time before the collector current starts to fall.

CHAPTER 7

Integrated Circuits

NO MODERN book on electronics would be complete without reference to integrated circuits. Although obviously related to established semiconductor technology, they represent a major revolution in electronics that has, or soon will have, the effect of making much present equipment obsolete. The depth of the subject is such that the treatment here will be relatively brief.

7.1. Types of Integrated Circuit

The term integrated circuit has been used to cover thin film, thick film, monolithic, and hybrid circuits. Of these the monolithic type is the most popular and has considerable advantages over the other types, but hybrid circuits—combining thin films and monolithic circuits as well as discrete components—are also widely used.

Monolithic circuits themselves have all the components and connecting leads manufactured on or near the surface of very small (0·06 by 0·06 in.) silicon blocks or dice. Thin film circuits have passive components manufactured by deposition or sputtering on a ceramic substrate with discrete active components mounted on top and connected by metalized interconnections. The thickness of thin film circuits is usually assumed to lie between 5 Å (1 Å = 10^{-8} cm) and 1 μ (10^{-4} cm), while above this they are termed thick films. In this chapter attention will be focused on the monolithic type.

7.2. Integrated Circuit Production Processes

Before considering the integrated circuits themselves, it is worth while discussing the manufacturing processes involved. A typical flow diagram for an integrated circuit is shown in Fig. 7.1, and this indicates the various processes required.

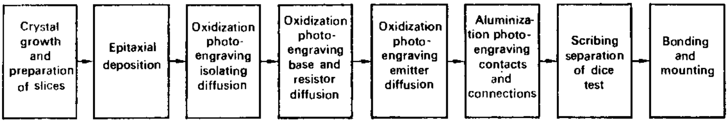

Crystal growth and preparation of slices
Epitaxial deposition
Oxidization photo-engraving isolating diffusion
Oxidization photo-engraving base and resistor diffusion
Oxidization photo-engraving emitter diffusion
Aluminization photo-engraving contacts and connections
Scribing separation of dice test
Bonding and mounting

FIG. 7.1. Flow diagram of integrated circuit.

7.2.1. *Oxidization*

The importance of a silicon dioxide layer first became apparent with the manufacture of planar and epitaxial planar transistors where the very low leakage currents are a direct result of the junctions being formed under a silicon dioxide layer. During the manufacture of integrated circuits, the surface is oxidized two or three times, and after selective etching provides a mask for the diffusion of impurity atoms or fabrication of connections.

One method of oxidizing the surface of a silicon wafer is to pass oxygen over it at about 1200°C. The thickness of the oxidized layer depends on the time and the temperature, and is typically about 5000 Å thick. A faster growth may be obtained by passing steam over the wafer, but even under these conditions a layer about 50 by 10^{-6} in. thick takes about an hour to be formed.

7.2.2. *Photo-engraving*

In the previous section reference is made to the etching of the silicon dioxide layer on the surface of the silicon wafer. This is a photo-etching process and involves two distinct operations—the preparation of the photographic mask and the etching of the silicon dioxide to cut the windows for the diffusion stages.

(a) Mask production

The final photographic mask comprises a large number of identical elements, each one a reduced version of an original artwork or drawing. Due to limitations in photographic equipment, several stages of reduction are required, and a typical series of stages is preparation of artwork originals, photographing, step-by-step contact printing, and re-photographing.

The usual method of preparing the original artwork is to cut out the required pattern in rubylith. This has a red top layer which when cut away reveals a white or clear mylar backing. This is suitable for most purposes, but, if greater stability is required, gelatin, dyed orange, is removed from a glass surface. If a drawing is used as the original, care must be taken to choose a highly reflecting paper which is dimensionally stable and accepts ink without spreading.

The pattern is then illuminated, and by conventional photography a negative, that gives a reduction of between ten and forty times, is obtained. Depending on whether a white or translucent backing is used, front or rear illumination is employed, the latter in practice giving a better edge definition. Exposure times are very critical since under-exposure may produce pinholes in the opaque areas while over-exposure may affect the degree of transparency in the clear areas.

The next stage is to mount the negative very accurately and by a series of contact prints produce an array of images. This step-by-step process must obviously be precisely controlled. The final stage is to re-photograph these multiple images, again giving a reduction of between ten and forty times. The resulting mask would typically have 200 or more images on a circle of $1\frac{1}{4}$ in. diameter.

(b) Photo-etching

With the photographic mask prepared, the processing of the slice may be started. Initially the surface of the slice is covered with a thin film of photo resist. This may be painted on, but a better method is to rotate the slice at about 8000 rev/min and drop some photo resist on the surface. Centrifugal force then spreads the light-sensitive emulsion evenly over the slice.

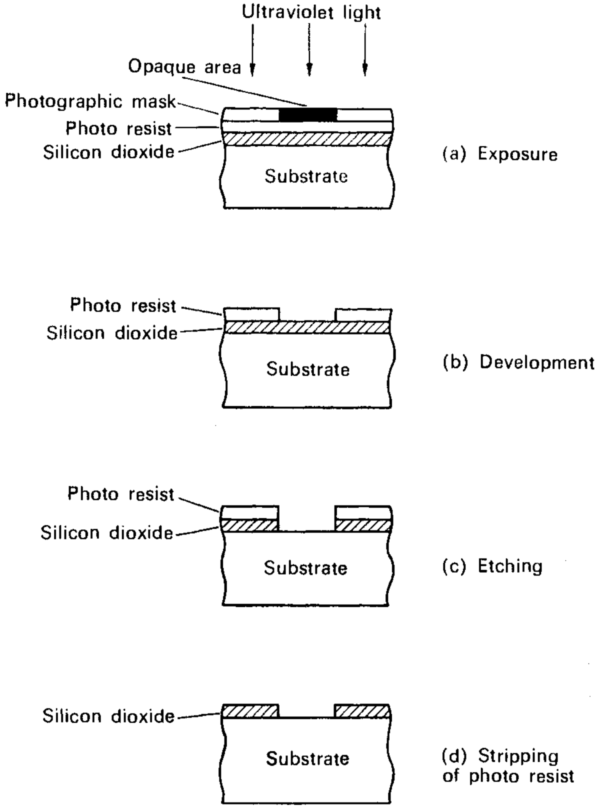


FIG. 7.2. Photo-etching.

After drying the photo resist in an oven, the photographic mask, previously prepared, is placed in the exact position required and the slice exposed to ultraviolet light. The opaque areas of the mask correspond to the areas to be etched away, and the subsequent action of the developer is such that it removes the photo resist in these areas. The other areas that have been polymerized are untouched by the developer.

The polymerized photo resist forms a coating that is resistant to hydrofluoric acid, and the use of this etchant enables the bare silicon dioxide to be etched away. The stages needed to open the windows are illustrated by Fig. 7.2.

7.2.3. Diffusion

In the preparation of integrated circuits there is always at least three diffusion stages. In fact the complete circuit may be produced in this manner, but it is usual to use a combination of epitaxial deposition and diffusion. The diffusion processes take place through the holes etched in the silicon dioxide, and the choice of diffusant must be such that it diffuses easily into the silicon itself but not through the dioxide. Boron and phosphorus are the usual impurities since gallium, for example, diffuses almost equally well through the oxide as the silicon itself.

In general, plating or evaporation tends to damage the surface, and hence impurities are normally carried to the wafer, which is held at an elevated temperature, in a gas stream, e.g. phosphorus pentoxide or boron tribromide in nitrogen. The time and concentration of impurities may be accurately controlled, and hence diffusion depths may be realized that are very close to the design values. The one exceptation to this is when gold is diffused, in which case the element is initially evaporated on to the surface.

7.2.4. *Epitaxial deposition*

This process, originally revived for the manufacture of discrete transistors, finds wide use in the production of integrated circuits. Diffusion, by its very nature, leads to an exponential distribution of impurities, and it is common practice to grow an *n*-type epitaxial collector layer on a *p*-type substrate. Such a layer is an extension of the crystal structure of the substrate and has a uniform distribution of impurities. Typically its resistivity is about 0·5 Ω cm compared with 10 Ω cm for the *p*-type substrate.

To reduce the possibility of surface imperfections extending into the epitaxial layer, the surface of the substrate is normally initially etched with gaseous hydrochloric acid. The epitaxial layer is then grown by deposition of atoms from a gaseous mixture that is passed over the substrate, the latter being heated to about 1200°C. Typically, hydrogen and silicon tetrachloride deposits the intrinsic material with phosphine for additional *n*-type impurities. The process is illustrated by Fig. 7.3, and it is relatively easy to process ten slices, each with about 200 monolithic circuits, simultaneously.

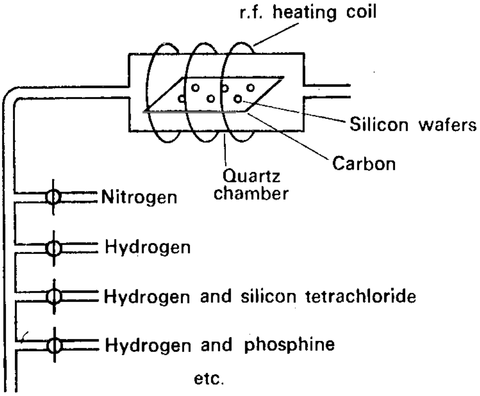


FIG. 7.3. Epitaxial deposition.

7.2.5. Evaporation

This process is normally used for the preparation of ohmic contacts and construction of interconnections in integrated circuits. Aluminium is the usual choice, but gold and nickel are also used. One difficulty is to avoid changing the nature of the semiconductor material, e.g. from *n* to *p*, by the addition of aluminium.

The process is a vacuum one, and after removing the oxide over the required contact areas, aluminium rods are evaporated producing a thin film over the entire surface. By masking and photoetching the undesired aluminium is removed to leave the required contacts and interconnections.

7.3. Active Circuit Components

In this section individual active components of integrated circuits will be considered. It is perhaps unfair to consider individual components in isolation, since all components on a slice are manufactured simultaneously, but later in the chapter the fabrication of a complete circuit will be discussed. It is no accident that active components are considered first since their number in any integrated circuit is considerably in excess of passive ones. In fact circuit design is such that the number of passive components, particularly large values, is kept to a minimum. Conventional diodes and bipolar transistors are the most important devices contructed, but other components such as field effect transistors are coming into increasing use.

7.3.1. Diodes

Due to the need for isolation between devices, in fact provided by reverse-biased *pn* junctions, *pn*-junction diodes in integrated circuits are actually three-layer devices as indicated by Fig. 7.4 (d). The stages in the construction are indicated by Fig. 7.4.

Figure 7.4 (a) shows the *p*-type substrate with a grown *n*-type epitaxial layer and a thin dioxide layer, while Fig. 7.4 (b) illustrates the die after windows have been etched and a *p*-type isolation diffusion performed; Fig. 7.4 (c) shows the diode structure after further etching and *p*-type diffusion. The final process is the aluminium evaporation which completes the diode as illustrated by Fig. 7.4 (d).

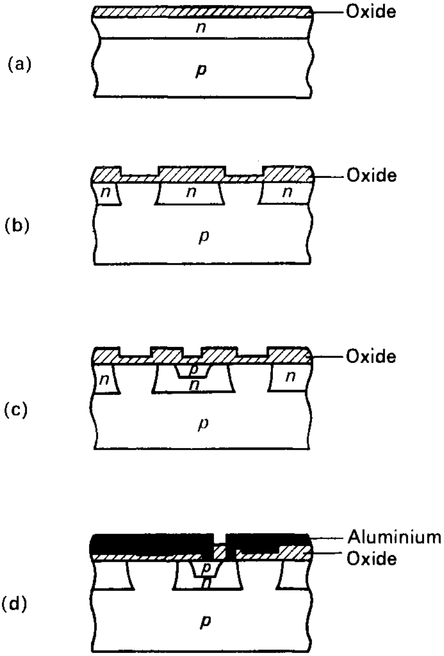


FIG. 7.4. Fabrication of a diode in an integrated circuit.

The need for three layers means that the equivalent circuit is more complex than that of a conventional diode, and one representation is shown in Fig. 7.5. In this, each diode has the usual limitations imposed by leakage conductance and junction capacitance, and these parameters should be very low. Design is, however, complicated by the fact that a low junction area, required for low capacitance, will mean a higher series resistance.

It is possible to fabricate other diodes, e.g. tunnel and backward diodes, on silicon dice, but these would probably require a separate diffusion stage due to the high impurity content of such diodes.

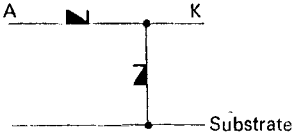


FIG. 7.5. Equivalent circuit of a diode in an integrated circuit.

The backward diode has not been mentioned previously in these volumes, and is one with a high impurity content—just less than that of a tunnel diode. Due to the very low reverse voltage breakdown it exhibits better rectification properties in the reverse than the forward direction.

7.3.2. Transistors

Since a diode in an integrated circuit is effectively a three-layer device, it follows that a transistor becomes a four-layer device. The system in which the collector is epitaxially grown is the most popular, but diffused collector and triple diffusion methods are used. As indicated by the following subsections, *npn* devices are much more popular than *pnp* ones.

(a) Diffused collector transistors

Figure 7.6 illustrates the construction of a diffused collector transistor in an integrated circuit; Fig. 7.6 (a) shows the substrate after oxidization, etching of windows, and *n*-type diffusion. Figure 7.6 (b) shows a cross-section after the base diffusion and Fig. 7.6 (c) the result of a third (emitter) diffusion. Figure 7.6 (d) shows the complete device with aluminium contacts and connecting leads.

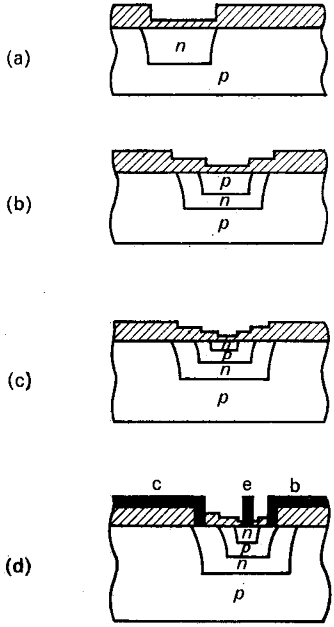


FIG. 7.6. Fabrication of a diffused-collector transistor.

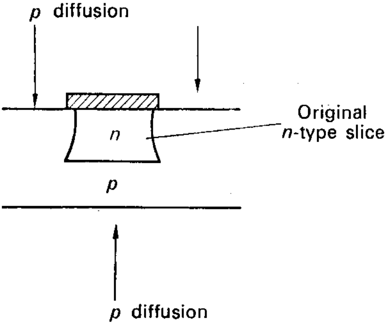


FIG. 7.7. Fabrication of a triple-diffused transistor.

(b) Triple-diffused transistors

When this type of transistor is incorporated in an integrated circuit, the original substrate is *n*-type and initially *p*-type impurities are diffused into both sides of the slice. On one side impurities are diffused into the whole slice to change it to a *p*-type substrate, while a mask on the other side enables an isolated *n*-type region to be fabricated as shown in Fig. 7.7. The *n* region has a uniform distribution of impurities, and further *p*- and *n*-type diffusing, as in the previous subsection, enable the complete transistor to be fabricated.

(c) Epitaxial grown and diffused transistor

The collector for this type of transistor is an *n*-type epitaxial-grown layer on a *p*-type substrate. This stage is shown in Fig.

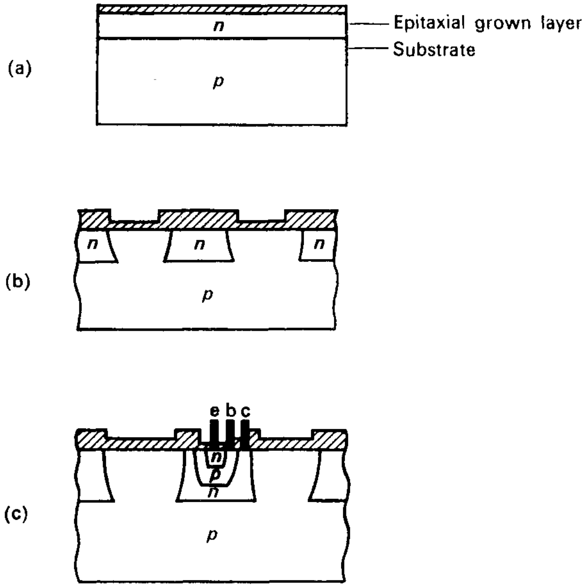


FIG. 7.8. Fabrication of an epitaxial transistor.

7.8 (a), while Fig. 7.8 (b) indicates the result of a *p*-type isolating diffusion. Further diffusions of *p* (base) and *n* (emitter) impurities and evaporation of contacts give rise to the structure illustrated by Fig. 7.8 (c). When the emitter is fabricated, the opportunity is often taken to diffuse extra *n*-type impurity into that region of the collector where the contact is made.

(d) Buried-n layer

In each of the three layers fabricated on the substrate, the added impurities must overcome that already present before changing the nature of the semiconductor. Apart from limiting the maximum number of diffusion stages to three, this means that the total impurity (donor and acceptor) increases from collector to emitter. Since the total impurity affects the mobility and hence the electrical characteristics, the doping of the collector itself must perforce be low. This fact, combined with the very small dimensions involved, gives rise to a high collector resistance. Its effect is often counteracted by a diffused low-resistivity layer known as a buried-*n* layer as shown in Fig. 7.9. The lines of current flow indicate the improvement possible by this method.

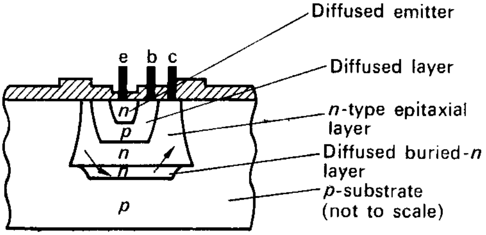


FIG. 7.9. The buried-*n* layer.

(e) Comparison of transistors made by different methods

Of the three types of transistor used in integrated circuits, the one fabricated using an epitaxial layer for the collector is the most popular. The difference arises from the different methods of construction used for the collector since these affect the concentration gradients and spread of impurities.

In the diffused-collector method the very nature of the diffusion process means that the greatest concentration of impurities is near the base–collector junction. This gives rise to a relatively high collector capacitance and low collector–base breakdown voltage. These difficulties do not arise with *pnp* devices if the substrate is used as the collector.

Since relatively long diffusion times are necessary to fabricate a triple-diffused transistor, the impurities will diffuse a relatively long distance under the oxide. This considerably reduces the useful area of the *n*-type structures, and the component density using this form of construction is much less than the other types.

The disadvantage of the above methods is largely overcome by using an epitaxial layer for the collector, and this type permits low collector capacitances with reasonable breakdown voltages to be achieved with high component densities.

(f) Equivalent circuit of transistor in an integrated circuit

As in the case of the diode, the isolating *pn* junction complicates the equivalent circuit of a transistor in an integrated circuit. It may be represented by Fig. 7.10, the diode representing the isolating junction. The capacitance of this diode does in fact limit the transient response of the transistor.

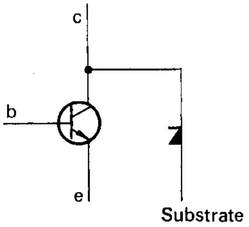


FIG. 7.10. Equivalent circuit of a transistor in an integrated circuit.

7.3.3. *Field effect transistors*

The construction of junction FETs follows a very similar pattern to junction bipolar transistors. If a *p*-channel device is considered, the stages in construction are the growth of an *n*-type epitaxial layer, *p*-type isolating diffusion, *p*-type channel diffusion, *n*-type gate diffusion, and, finally, evaporating of contacts. The complete device is shown in Fig. 7.11.

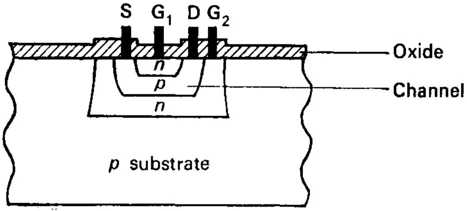


FIG. 7.11. Junction field effect transistor.

The insulated-gate FET promises much for the future and has the advantage that the number of separate stages required is reduced. If an *n*-channel enhancement type is considered, the stages are diffusion of *n*-type drain and source regions into a *p*-type substrate, oxidization, and evaporation of gate electrode and source and drain contacts. The complete device is illustrated by Fig. 7.12, and although it has a slower response than a bipolar transistor, it occupies about one-tenth of the area.

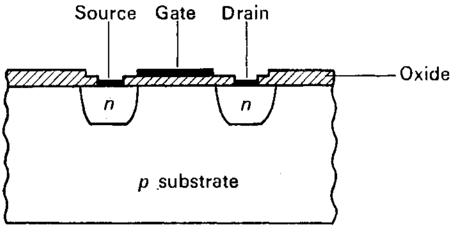


FIG. 7.12. MOST in an integrated circuit.

7.4. Passive Integrated Circuit Components

Passive components in integrated circuits include conductors, diffused and thin film resistors, and junction and thin film capacitors. As stated previously, the number of these is kept to a minimum, and in many cases active devices now perform the role previously performed by passive ones.

7.4.1. Conductors

Although interconnections are in general evaporated on the surface of the die, it is necessary in some instances to have crossovers. In such cases conductors must be fabricated in the silicon die itself. After oxidization, photo-etching, and *n*-type diffusion to produce the conductor, further oxidization, evaporation, and photo-etching enables a crossover to be constructed as indicated in Fig. 7.13.

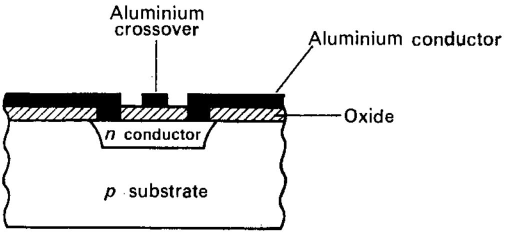


FIG. 7.13. Crossover in an integrated circuit.

7.4.2. Diffusion resistors

The construction of a diffused resistor is illustrated by Fig. 7.14. As in the case of conductors, initial *n*-type epitaxial-layer deposition and isolating diffusion are followed by *p*-type diffusion of the actual resistor and evaporation of contacts. In practice these stages occur at the same time as other devices are being fabricated.

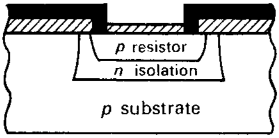


FIG. 7.14. Diffused resistor.

The resistance R of any resistor is given by the formula

$$R = \varrho\frac{l}{A},$$

where ϱ is the resistivity, l is the length, and A is the area.

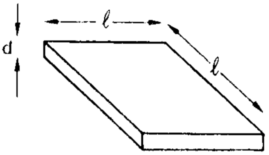


FIG. 7.15. Ohms per square.

If the length is made equal to the width as shown in Fig. 7.15, the area $A = ld$.

Therefore

$$R = \frac{\varrho}{d}.$$

Hence for a given value of d the resistance R of a square of material is independent of the length of the sides. The resistance may then be quoted in ohms per square. This is the notation used in integrated circuits, and, typically, values of 100–200 Ω per square are obtained with regions 10^{-4} in. thick.

The design of resistors in integrated circuits is based on the idea of a square, or rather a number of squares. One that is comprised of ten squares in line, i.e. its length being ten times its width, has a resistance of between 1000 and 2000 Ω. In practice a compromise must be reached between the area required and the accuracy of the photo-engraving process. This normally limits the maximum value of diffused resistors to about 25 kΩ.

Due to the presence of the isolation diffusion, the equivalent circuit of a diffused resistor is relatively complex, and one representation is shown in Fig. 7.16. At high frequencies the capacitance of the diodes may have deleterious results.

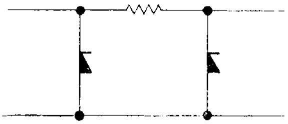


FIG. 7.16. Equivalent circuit of a diffused resistor in an integrated circuit.

7.4.3. Thin film resistors

The disadvantages of diffused resistors include their limited value, high temperature coefficient of resistance, susceptibility to radiation, and the fact that they cannot be adjusted individually. The last difficulty is, however, not as much trouble as might be anticipated since the tolerance between resistors deposited, at the same time, is probably better than 5%. The other disadvantages may be overcome by using thin film resistors.

Thin film resistors are usually produced by depositing nichrome or tantalum on top of the silicon dioxide. Tantalum has the advantage that it is a single element and not an alloy, and hence there is no danger of composition changes during fabrication. The method of deposition itself, may be either plating, evaporation,

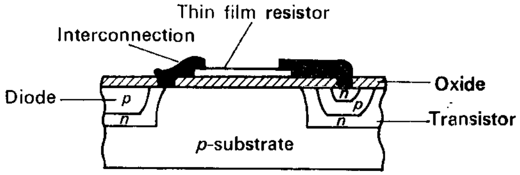


FIG. 7.17. Thin film resistor.

or sputtering. The last mentioned is simplest since lower temperatures are involved. A typical thin film resistor incorporated in a hybrid circuit is shown in Fig. 7.17.

7.4.4. *Active resistors*

The obvious choice, using bipolar transistors, for a high-impedance resistor is a bootstrap circuit as shown in Fig. 7.18. Due to feedback, the apparent input resistance is very high and the bias circuit is such that the reduction due to the d.c. base feed, occurring in conventional common-collector circuits, is nullified.

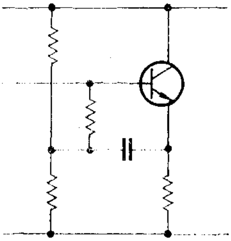


FIG. 7.18. Bootstrap circuit.

7.4.5. *Diffused capacitors*

This circuit element is provided by the capacitance of a *pn* junction and, as with diffused resistors, there is an upper limit to the component value that may be reasonably achieved. The construction follows the same pattern as diodes, but the impurity gradient, and hence the value of the capacitance per unit area, may

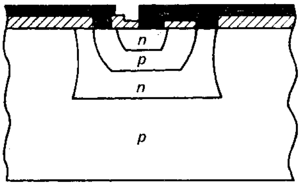


FIG. 7.19. Diffused capacitor.

be increased by using a buried-*n* layer. An area of 0·001 in^2 corresponds to about 2·5 pF, and capacitances of 1000 pF may be achieved, although normally it is limited to 300 or 400 pF.

The capacitor is illustrated by Fig. 7.19, while Fig. 7.20 illustrates an equivalent circuit. The shunt resistance is that of the reverse-biased diode, and its value is such that the effective *Q* factor is only about 10.

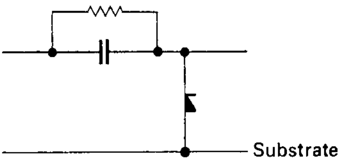


FIG. 7.20. Equivalent circuit of a diffused capacitor in an integrated circuit.

7.4.6. *Metal oxide capacitors*

Due to the low *Q* factor previously mentioned, the use of junction capacitors is somewhat restricted. A better capacitor may be produced by using the silicon dioxide as a dielectric as indicated in Fig. 7.21. The *Q* factor of this type is about ten times that of the junction capacitor, and better tolerances are obtainable.

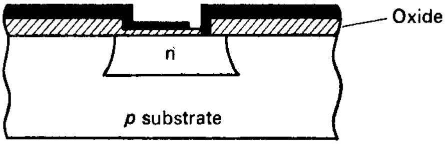


FIG. 7.21. Metal oxide capacitor.

7.4.7. *Active circuit substitutes for capacitors*

As mentioned previously, the design of integrated circuits is such that the number of passive components is kept to a minimum. Emitter-decoupling capacitors are not generally used since the loss of gain, due to omitting it, may easily be obtained by adding an extra stage. The resulting negative feedback gives advantageous results, and the circuit—known as a Darlington pair—is illustrated by Fig. 7.22.

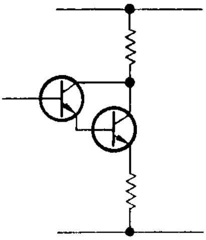


Fig. 7.22. Darlington pair.

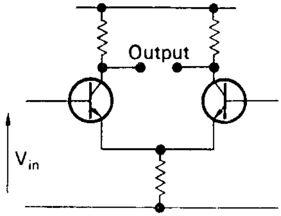


Fig. 7.23. Long-tail pair.

Coupling capacitors are also largely eliminated from integrated circuits by using long-tail pair configurations as illustrated by Fig. 7.23. In this case a balanced output is obtained from the two collectors.

7.4.8. *Active circuit substitutes for inductors*

Inductors themselves are not fabricated in integrated circuits, but several substitutes for inductance may be used. The action of a high-frequency choke, for instance, may be simulated by a field effect transistor (FET). Typical characteristics, as shown in Fig. 7.24, show a relatively low d.c. resistance but a very high a.c. resistance, while Fig. 7.25 indicates how this may be incorporated in an integrated circuit.

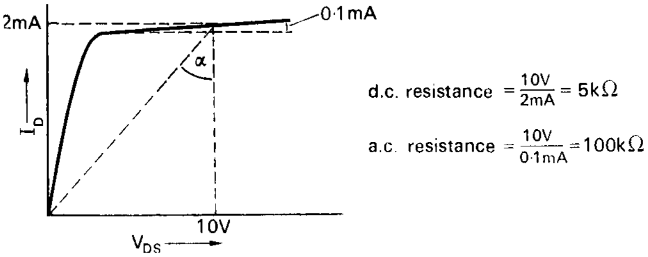


Fig. 7.24. Characteristics of a FET.

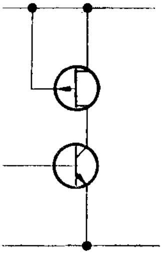


FIG. 7.25. Use of FET as a high impedance choke.

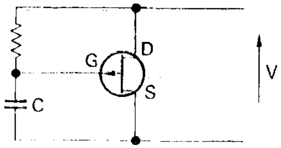


FIG. 7.26. Reactance FET circuit.

For tuning purposes RC networks may often be used instead of LC circuits, but if an inductance is essential a variation of the reactance valve circuit may be used. Figure 7.26 shows such a circuit using a FET that takes a lagging current. The FET may be represented by a constant current generator as shown in Fig. 7.27,

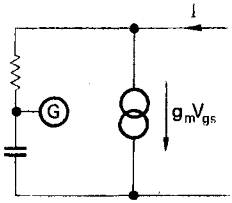


FIG. 7.27. Equivalent circuit of Fig. 7.26.

and to a close approximation an alternating voltage V applied between drain and source gives rise to a current g_mV_{gs}, where

$$V_{gs} = \frac{1/j\omega C}{R+(1/j\omega C)}V = \frac{V}{1+j\omega CR}.$$

Hence the impedance between drain and source

$$Z_{ds} \triangleq \frac{V}{g_mV/(1+j\omega CR)} = \frac{1+j\omega CR}{g_m}$$
$$= \frac{1}{g_m}+j\frac{\omega CR}{g_m}.$$

This represents a resistance of $1/g_m$ in series with an inductance of CR/g_m.

7.5. Realization of an Integrated Circuit

Previously in this chapter the processes involved and the details of individual components have been considered. Although necessary, this tends to give an incomplete picture since the whole point of integrated circuit production is the simultaneous fabrication of a large number of components. In this section an indication of the complete manufacturing process will be given.

The designer of integrated circuits must think initially in terms of circuit functions and how they may be realized. This is a different approach from that of discrete circuit design where initial thoughts are usually in terms of device characteristics and how they may be utilized. The restrictions on circuit elements have already been indicated, and they are that passive components are kept to a minimum, inductances are not used, and the number of active devices is relatively unimportant.

To illustrate the manufacturing processes involved, consider the part of a circuit shown in Fig. 7.28. This is not necessarily any particular circuit but is more a collection of components.

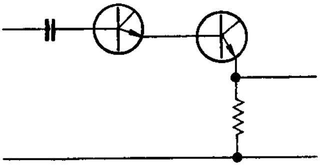


FIG. 7.28. Part of typical integrated circuit.

The first stage in the fabrication of the integrated circuit is the production of the *p*-type silicon slice. Initially the crystal must be grown, and this follows the well-established pattern for discrete devices. Chemical and metallurgical methods reduce the impurities to a low level, and subsequent zone-refining results in an impurity level of better than 1 part in 10^{10}. In the latter a long ingot of semiconductor material is drawn through a series of radio-frequency coils so that a molten zone travels along the ingot. The impurities travel with the molton zone to the end of the ingot, leaving the majority of the ingot free of impurity. This pure semiconductor material is then melted with a known amount of impurity, and a single crystal drawn from the melt. The resulting crystal would be about 6–8 in. long and 1 in. in diameter. It is then sliced and polished to a smooth finish. The slice about 0·005 in. thick is then ready for the next stage in the fabrication of the integrated circuit.

At the same time as—or before—the production of the slice, the photographic masks would be prepared. Assuming the final structure as shown in Fig. 7.33 (p. 208), this would include masks for:

(a) The buried-*n* layer.
(b) *p*-Type isolating diffusion.
(c) *p*-Type capacitor region, *p*-type resistor, and *p*-type base region.
(d) *n*-Type capacitor region and *n*-type emitter regions.
(e) Aluminium connections.

The first diffusion stage is that required to produce the buried-*n* layer. After oxidization of the *p*-type slice, the first mask is used to cut windows and an *n*-type diffusion of antimony or arsenic atoms performed. These are relatively slow diffusers, since during subsequent operations there is a further spread of impurities. The part of the die considered is represented by Fig. 7.29 at this stage. Due to the high impurity content, this layer is designated *n*+.

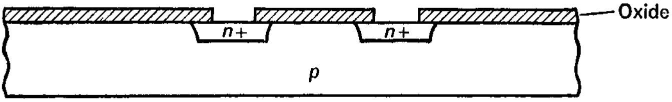


FIG. 7.29. Fabrication of Fig. 7.28—stage 1.

Removal of the oxide followed by epitaxial deposition of an *n*-type layer is the next stage, and this continues until the layer is about 0·0005 in. thick. Allowing for lateral as well as vertical spreading of the isolating diffusion, this thickness enables an optimum component density to be achieved. After reoxidization, the part of the slice is now as shown in Fig. 7.30.

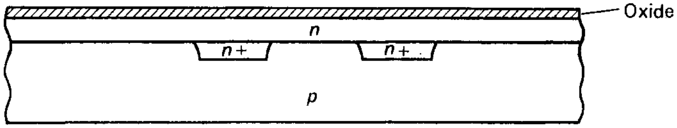


FIG. 7.30. Fabrication of Fig. 7.28—stage 2.

After covering with photo resist, the second mask is now placed exactly over the slice containing, perhaps, 200 separate circuits. Exposure and development follow to enable windows for the isolating diffusion to be etched in the oxide. Boron would probably be used for this subsequent diffusion, giving a structure after reoxidization as indicated in Fig. 7.31.

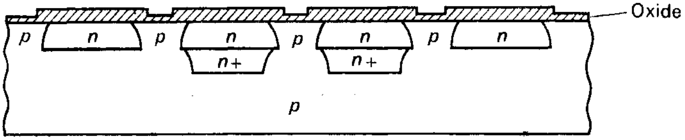


FIG. 7.31. Fabrication of Fig. 7.28—stage 3.

A similar series of photo-engraving stages and boron diffusions would enable the *p*-type region of the junction capacitor and diffused resistor and the base regions of the transistors to be fabricated. This is followed by photo-engraving and diffusion of the *n*-type region of the capacitor and the *n*-type emitter. Phosphorus, which is a more rapid diffuser than arsenic, would be used for these diffusions. The part of the slice being considered would, after oxidization, be as indicated in Fig. 7.32.

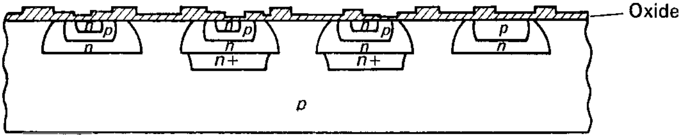
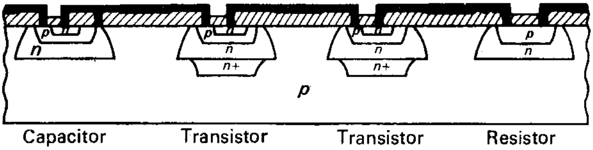


FIG. 7.32. Fabrication of Fig. 7.28—stage 4.

The diffusion stages would be finished at this time, but a further photo-engraving process is required to open up windows in the oxide to allow contacts to be made to the devices. Vacuum deposition of aluminium over the entire surface and a final photo-etching process to remove the unwanted aluminium would complete the fabrication of the slice itself, at this stage comprising, perhaps, 200 untested circuits. The part previously considered is shown in Fig. 7.33 and from this stage the more expensive processing of the individual dice would be required. As an indication of the dimensions involved, the separation of elements is about 0·0005 in. and the thickness of the *p*-type resistor is about 0·0001 in. Its width is of the same order.



Note: collector leads not shown

FIG. 7.33. Realization of Fig. 7.28.

There remains the problem of testing and separation of the individual dice. Initial testing using multi-electrode probes are made on the complete slice, and the faulty dice are marked with ink. Diamond scribing and separation follows, and the faulty circuits are discarded.

The dice, about 0·06 in. square, would then be bonded onto the gold-plated encapsulation. A simple procedure for this is to raise the temperature of the die and encapsulation to about 380°C when a gold-silicon eutectic is formed between the slice and the encapsulation. On cooling, the eutectic solidifies and the die is held firmly in place. The encapsulation itself is usually one of two forms, TO5 can or flat pack, and are illustrated with the fine-wire (usually 0·001 in. diameter aluminium) connections by Fig. 7.34 (a) and (b). The final production stage is to seal the unit, resistance welding for the TO5 can or soldering for the flat pack and, after final test as a complete unit, it is ready for use.

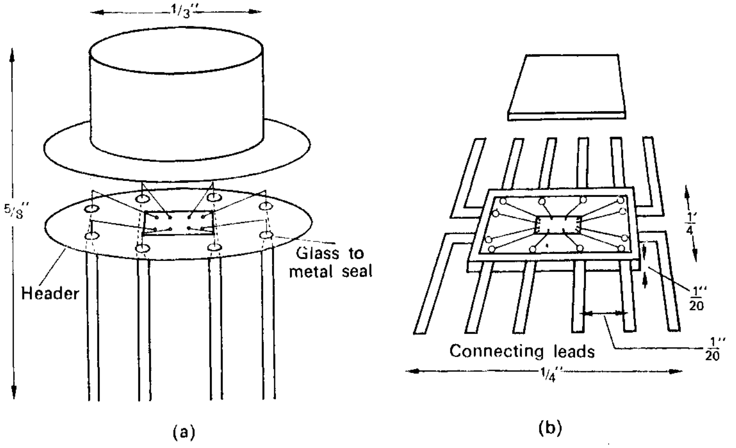


FIG. 7.34. Encapsulation of an integrated circuit.

Questions on Chapter 7

Q.1. Describe the manufecturing processes involved in the production of integrated circuits. Explain how these are used in the fabrication of one epitaxial transistor in a complete circuit.

Q.2. Discuss and compare the different methods available for the fabrication of bipolar transistors in integrated circuits.

Q.3. Outline the different types of passive component, or their equivalents, used in integrated circuits.

CHAPTER 8

Noise

IN THE early days of the thermionic triode it was realized that amplification would be limited by noise. In this chapter the various sources of noise will be outlined and their influence on amplifier performance considered.

8.1. Thermal Noise

This form of noise is associated with the flow of current through a resistance. Such a current is due to the drift of a large number of electron charge carriers which at any instant are not evenly distributed throughout the resistor and give rise to minute variations in potential between the terminations. Over a long period these variations will add up to zero, but, as far as the a.c. output is concerned, represent a noise voltage.

Thermal noise is not frequency dependent and is limited only by the system considered. It is, in fact, proportional to the noise bandwidth δf which, in general, is slightly different from the signal bandwidth. Since it is thermal in origin (i.e. dependent on the agitation of individual molecules), it is also proportional to the absolute temperature T, and the noise power developed in a resistor P_n is given by

$$P_n = kT\delta f \text{ watts.}$$

The constant k in this equation is, in fact, Boltzmann's constant, i.e. $1{\cdot}38 \times 10^{-23}$ J/K.

This is the noise power the resistor can deliver to a matched load. Under such conditions the noise current flowing I_n is given by

$$I_n^2 R = kT\,\delta f.$$

But due to the matching resistor the noise voltage generated in the resistor E_n is given by

$$E_n = I_n 2R.$$

Hence
$$\left(\frac{E_n}{2R}\right)^2 R = kT\,\delta f$$

or
$$E_n^2 = 4kTR\,\delta f.$$

This well-known relation gives the noise voltage generated in a resistor, and is valid up to about 10,000 MHz when quantum effects have to be taken into account.

Thermal noise is also known as Johnson, resistance, or fundamental noise, and is of major importance in the first stage of an amplifier. In valve circuits the critical resistance is the grid leak. Field effect transistor circuits have two main sources of thermal noise—the resistance associated with the first gate and the resistance of the conducting channel itself. For bipolar transistors, their lower circuit impedances tend to make the base resistance of the first transistor itself the main source of thermal noise.

It has already been stated that the noise bandwidth is normally slightly different from the signal bandwidth. When a circuit has a rectangular response as indicated in Fig. 8.1, the signal and noise bandwidths are identical, but a response curve as shown in Fig.

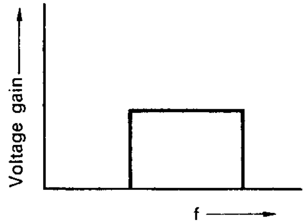


FIG. 8.1. Response curve for identical signal and noise bandwidth.

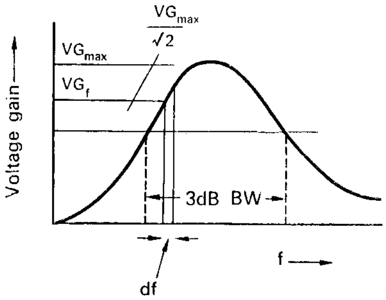


FIG. 8.2. General response curve.

8.2 leads to a slightly larger noise bandwidth. In the latter case, to include all the noise contributions the noise bandwidth must be defined as

$$\delta f = \int_0^\infty \frac{PG_f\, df}{PG_{\max}},$$

where PG_f is the power gain over a narrow band df and $PG_{\max}$ is the maximum power gain,

or $$\delta f = \int_0^\infty \frac{VG_f^2\, df}{VG_{\max}^2},$$

where VG_f is the voltage gain over a narrow band df and $VG_{\max}$ is the maximum voltage gain.

For many calculations, however, the difference between the two bandwidths may be ignored and the signal bandwidth is sufficiently accurate.

Norton's theorem enables the noise-voltage generator to be replaced by a noise-current generator. This is sometimes a useful conversion and is shown diagrammatically in Fig. 8.3, the voltage generator $\sqrt{(4kTR\ \delta f)}$ in series with a resistance R being converted into a current generator $\sqrt{(4kTG\ \delta f)}$ in parallel with a conductance G, where $G = 1/R$.

The noise generated in a resistor is purely random, and when two resistors are connected in series there is no correlation between

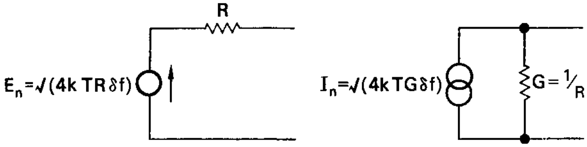


FIG. 8.3. Voltage- and current-noise generators.

the noise generated by each resistor. This means that the noise-generated by one is unlikely to peak at the same time as that generated by the other (or the reverse to occur that the noise developed by one cancels that due to the other).

For two resistors R_1 and R_2 in series,

$$\begin{aligned} E_n^2 &= 4kT(R_1+R_2)\,\delta f \\ &= 4kTR_1\,\delta f+4kTR_2\,\delta f \\ &= E_{n_1}^2+E_{n_2}^2. \end{aligned}$$

This equation shows that the squares of the noise voltages due to the individual resistors must be added and not the noise voltages themselves. This, in fact, is general for all random (uncorrelated) noise voltages.

When reactive circuits are considered, the value of R in the expression $E_n^2 = 4kTR\,\delta f$ is the real part of the impedance at the centre frequency. This may be illustrated by the tuned circuit shown in Fig. 8.4. The circuit itself may be assumed to be connected across the input of an amplifier.

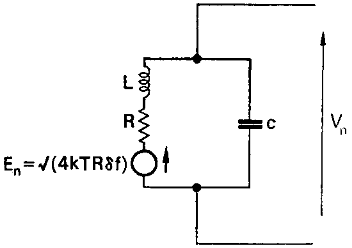


FIG. 8.4. Noise voltage developed by tuned circuit.

The resistance will generate a noise voltage and at the centre frequency the noise voltage across the capacitor V_n is given by

$$
\begin{aligned}
V_n &= \frac{\sqrt{(4kTR\,\delta f)}}{R}\,\frac{1}{\omega_0 C} \\
&= \sqrt{\left(\frac{4kT\,\delta f}{R}\,\frac{LC}{C^2}\right)} \quad \text{since} \quad \omega_0^2 = \frac{1}{LC} \\
&= \sqrt{\left[4kT\left(\frac{L}{CR}\right)\delta f\right]} \\
&= \sqrt{(4kTR_D\,\delta f)},
\end{aligned}
$$

where $R_D = L/CR$ is the dynamic impedance of the tuned circuit.

8.2. Shot Noise

Shot noise occurs when there is a random movement of charge carriers in and out of a region, normally depleted of charge carriers. The minute variations represent a noise current which depends on the mean d.c. current I flowing and, as in the case of thermal noise, on the noise bandwidth δf. It may be shown by assuming a Gaussian distribution of velocities that such a noise current is given by

$$I_n^2 = 2Ie\,\delta f, \quad \text{where } e \text{ is the charge on an electron.}$$

In valves, shot noise has the greatest effect in temperature-limited devices. Shot noise in temperature-limited diodes may be used, as shown later in this chapter, as the basis of a method of measuring the noise factor of amplifiers. When a space charge is present, some of the noise fluctuations are smoothed out, and under such conditions the noise current is given by $I_n = F\sqrt{(2Ie\,\delta f)}$, where F is a factor of 1 for temperature-limited conditions but decreases as the space charge becomes more effective.

In bipolar transistors shot noise is generated in the depletion layers, its value being given by $I_n = \sqrt{(2I_E e\,\delta f)}$ in the emitter–base diode and $I_n = \sqrt{(2I_C e\,\delta f)}$ or $\sqrt{(2I_E\,|h_{fb}|\,e\,\delta f)}$ in the collector–base diode. Shot noise is usually negligible in FETs, the only source being due to the gate leakage current, i.e. $I_n = \sqrt{(2I_G e\,\delta f)}$, where the gate current I_G is very small.

8.3. Flicker Noise (1/*f* Noise)

This form of noise occurs in both valves and transistors and, as its name implies, is only important at relatively low frequencies.

In most valve applications, flicker effect is negligible since it is small compared with shot noise. For tungsten cathodes for instance, it is only important below 100 Hz. It is caused by irregularities in cathode emission due to charges on the surface arising from diffusion or ion bombardment.

In bipolar transistors the main source of $1/f$ noise is crystal imperfections normally at the surface. Carriers diffuse in and out of traps or recombination centres giving rise to a noise current that rapidly increases at lower frequencies. This type of noise predominated in the first transistors produced, and even now $1/f$ noise is an important consideration in the choice of transistors for a particular application. High-frequency devices have a greater $1/f$ noise component than those which cut off at lower frequencies, and in general, for a high-gain amplifier it is not advisable to choose a transistor with a cut-off frequency much greater than that required.

Field-effect transistors have traps and recombination centres in the same way as bipolar transistors. Flicker noise is, hence, a source of noise in the FET and, as in the previous case, predominates at very low frequencies.

8.4. Other Noise Sources

The three types of noise already mentioned are the most important in electronic circuits, but other, normally minor, forms of noise become important under certain conditions.

In vacuum devices, perfect vacuums are not attainable, and there are always some residual gas atoms left after evacuation. The movement of ions formed from these atoms partly neutralizes the space charge and causes minute variations in anode current, i.e. ion noise. Also at very high frequencies a fluctuation in grid current is induced by the passage of electrons. This in turn modulates the electron beam giving rise to induced grid noise.

Multi-grid valves have two further forms of noise inherent in their operation. One is partition noise due to variations in the division of current between positively biased electrodes, and the other is secondary emission noise due to variations in secondary emission from positively biased electrodes. The latter is particularly important in photo multipliers.

Although they are forms of shot and/or thermal noise, other terms have been used to describe the noise in transistors. These are generation noise due to fluctuations in the generation of carriers (which is really shot noise), diffusion noise due to fluctuations in diffusion rates (which is really thermal noise), and modulation noise due to surface field effects.

In addition to noise sources inherent in devices, other forms can arise due to bad design and layout. These include microphony, poor contacts, hum due to electrostatic or electromagnetic coupling or ripple on the supply, and radiation or mains-borne noise.

8.5. Noise in Devices

In this section the noise in different electronic devices will be examined and, in the case of the diode, the effect of the diode load considered.

8.5.1. Noise in thermionic diodes

In a temperature-limited diode the shot noise is, as already stated, given by $I_n = \sqrt{(2Ie\,\delta f)}$. When, however, the diode load is added as shown in Fig. 8.5, the thermal noise generated in the resistor

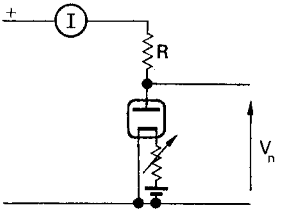


FIG. 8.5. Noise output of diode and load resistor.

R must be allowed for.

Then

$$V_n^2 = 4kTR\,\delta f + 2Ie\,\delta f R^2$$

or $$V_n = \sqrt{[2R\,\delta f(2kT + IeR)]}.$$

Under space-charge conditions, besides introducing a factor F in the expression for shot noise, the r_a of the valve must be allowed for. The noise equivalent circuit for such a combination is

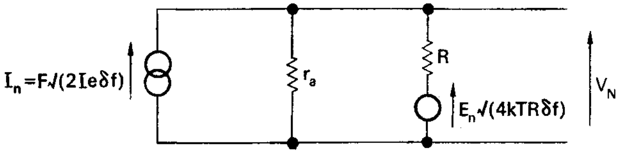


FIG. 8.6. Noise output of space-charge limited diode.

then as shown in Fig. 8.6. Hence, using the superposition theorem,

$$\begin{aligned} V_n^2 &= I_n^2\left(\frac{r_aR}{r_a+R}\right)^2 + E_n^2\left(\frac{r_a}{r_a+R}\right)^2 \\ &= 2F^2Ie\,\delta f\left(\frac{r_aR}{r_a+R}\right)^2 + 4kTR\,\delta f\left(\frac{r_a}{r_a+R}\right)^2 \\ &= 2\left(\frac{r_a}{r_a+R}\right)^2 (F^2IeR + 2kT)\,R\delta f \end{aligned}$$

or $$V_n = \left(\frac{r_a}{r_a+R}\right)\sqrt{[2(F^2IeR + 2kT)R\,\delta f]}.$$

8.5.2. Noise in triodes

The major source of noise in triode valves is shot noise due to variations in cathode emission. This is often referred to the grid where the effect of shot noise is equated to the thermal noise which would be produced by an imaginary resistor connected in series with the grid. This imaginary resistor is known as the equivalent grid resistor R_n. Such a resistor gives rise to a fluctuating anode current of $g_m \sqrt{(4kTR_n \, \delta f)}$ and equating this to $F\sqrt{(2Ie \, \delta f)}$,

$$g_m \sqrt{(4kTR_n \, \delta f)} = F\sqrt{(2Ie \, \delta f)}$$

or

$$R_n = \frac{F^2 Ie}{2kTg_m^2}.$$

An approximation often used is $R_n = \dfrac{2 \cdot 5}{g_m}\ \Omega.$

8.5.3. Noise in tetrodes and pentodes

As in the case of triode valves, shot noise is the major form of noise in multi-grid valves, but some addition to the equivalent grid resistance is required to allow for partition noise. The approximation for the equivalent grid resistance is often given by

$$R_n = \frac{I_A}{I_A + I_S}\left(\frac{2 \cdot 5}{g_m} + \frac{20 I_S}{g_m^2}\right),$$

where I_A and I_S are the d.c. values of anode and screen current respectively.

8.5.4. Noise in bipolar transistors

As has already been stated, over the normal range of operation the main sources of noise in transistors are shot noise in the emitter–base and collector–base depletion layers and thermal noise in the base spreading resistance. This is most conveniently shown by means of the "T" equivalent circuit. Such a circuit, including noise generators, is shown in Fig. 8.7, and is normally valid from about 1 kHz, i.e. above the range in which $1/f$ noise is appreciable, to the common-base (α) cut-off frequency.

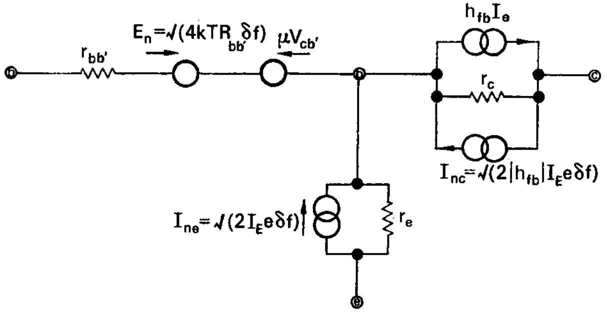

$r_{bb'}$
$E_n=\surd(4kTR_{bb'}\delta f)$
$\mu V_{cb'}$
$h_{fb}I_e$
r_c
$I_{nc}=\surd(2|h_{fb}|I_E e\delta f)$
$I_{ne}=\surd(2I_E e\delta f)$
r_e
b
b'
c
e

FIG. 8.7. Noise equivalent circuit of bipolar transistor.

The complete noise-frequency relation is shown in Fig. 8.8, and includes the low-frequency $1/f$ noise and the high-frequency frequency-dependent shot noise.

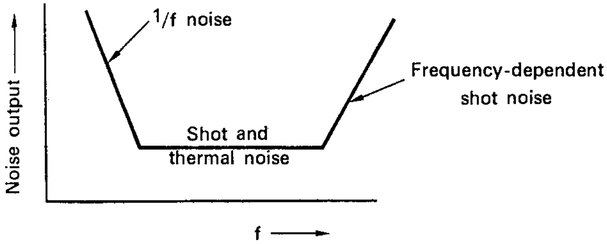


FIG. 8.8. Noise-frequency distribution of bipolar transistor.

8.5.5. *Noise in the FET*

These are inherently much lower noise devices than bipolar transistors, and although there is a small shot noise due to the gate current, normally the only appreciable noise generated is the thermal noise developed in the channel itself. An expression given for this is $4kT\,\delta f/g_m$, where g_m is the mutual conductance, and indicates that a FET should be operated at the highest possible value of g_m, which in turn implies a high drain current. Incidentally, $1/f$ noise in field-effect devices may be regarded as a form of thermal noise that increases rapidly at very low frequencies.

8.6. Noise in Complete Amplifiers (i.e. Black-box Representation)

Noisy amplifiers may be represented by noiseless amplifiers with noise source(s) connected across the input terminals. It is common practice to use different models for valve and transistor amplifiers, even though the valve model may be regarded as a simplification of the transistor model.

8.6.1. *Equivalent noise source for a valve amplifier*

The idea of an equivalent noise resistor has already been introduced in section 8.5.2 and may be extended to a complete valve amplifier. If, for the amplifier shown in Fig. 8.9, the output noise, measured by a true r.m.s. instrument, is plotted for different values of resistance R, the relation will be as shown in Fig. 8.10. This shows the output noise power being due to thermal noise power generated in the resistor R plus internal noise power generated within the amplifier. The internal noise, as shown by the intercept on the vertical axis (i.e. corresponding to $R = 0$), is known as the

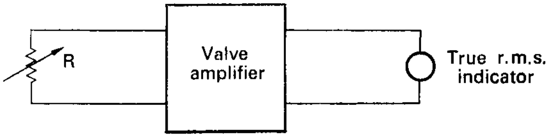


FIG. 8.9. Measurement of equivalent noise resistance.

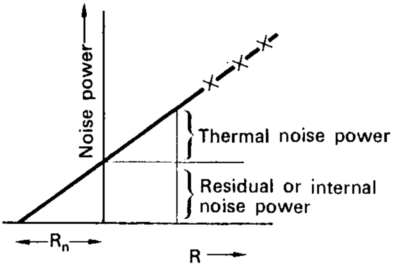


FIG. 8.10. Determination of equivalent noise resistance.

residual noise power and may be regarded as being produced by a resistor R_n in series with the input.

8.6.2. *Equivalent noise sources for a transistor amplifier*

Transistors have several sources of noise inherent in their operation, and to allow for these the black-box representation of a noisy transistor amplifier is as shown in Fig. 8.11. This shows a series-voltage noise generator E_n and a parallel-current noise generator I_n with a correlation factor γ between them.

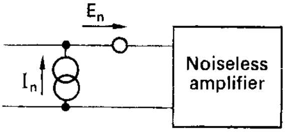


FIG. 8.11. Equivalent noise sources of a transistor amplifier.

The voltage generator is equivalent to a noise resistor R_{n_1}, where $R_{n_1} = \dfrac{E_n^2}{4kT\,\delta f}$ and may be determined by short circuiting the input, measuring the output noise and referring it back to the input. This corresponds to R_n, the noise resistance discussed in the previous section. The current generator may also be replaced by a resistor R_{n_2}, where $R_{n_2} = 4kT\,\delta f/I_n^2$, and may be determined by open circuiting the input, measuring the output noise and referring it back to the input. To evaluate γ, a third noise measurement must be made, typically with a source resistance of $\sqrt{(R_{n_1} R_{n_2})}$. The noise equivalent circuit is then as shown in Fig. 8.12, but care must be taken with its use as shown later in this chapter.

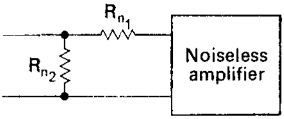


FIG. 8.12. Equivalent noise resistances of a transistor amplifier.

8.7. Signal-to-noise Ratio

When intelligibility is the criterion, e.g. in radio receivers, the signal-to-noise (S/N) ratio of a system is often much more important than the sensitivity. It is useless to have such a high gain that the input signal is masked by noise, and the sensitivity is often stated in terms of the signal required for a certain S/N ratio (say 15 dB) at the output. The S/N ratio itself is always given as a power ratio.

In the simplest case, i.e. an infinite input impedance noiseless amplifier, the only source of noise is the thermal noise associated with the source resistance R_S. If the input signal is E, the ratio of signal power to noise power at the output is $E^2/4kTR_S\,\delta f$, where the symbols have the meanings previously defined.

When the input resistance has a finite value, but no other noise is generated within the amplifier, the equivalent circuit is as illustrated by Fig. 8.13. This shows the input voltage is reduced to

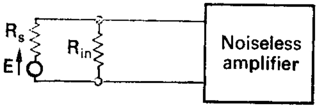


FIG. 8.13. S/N ratio of matched amplifier.

$ER_{\text{in}}/(R_S+R_{\text{in}})$ while the noise voltage E_n becomes

$$\sqrt{\left[4kT\left(\frac{R_S R_{\text{in}}}{R_S+R_{\text{in}}}\right)\delta f\right]}.$$

The S/N ratio at the output is then given by

$$\text{S/N ratio} = \frac{E^2R_{\text{in}}^2/(R_S+R_{\text{in}})^2}{4kT[R_SR_{\text{in}}/(R_S+R_{\text{in}})]\delta f}$$

$$= \frac{E^2R_{\text{in}}}{4kT(R_S+R_{\text{in}})R_S\,\delta f}.$$

The inclusion of a noise resistance R_n in series with the input as

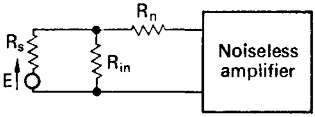


FIG. 8.14. S/N ratio of matched amplifier with noise resistance.

illustrated by Fig. 8.14 has no effect on the signal but increases the effective noise resistance to $\left(\frac{R_SR_{\text{in}}}{R_S+R_{\text{in}}}+R_n\right)$,

$$\text{i.e.}\quad \text{S/N ratio} = \frac{E^2R_{\text{in}}^2/(R_S+R_{\text{in}})^2}{4kT\left(\frac{R_SR_{\text{in}}}{R_S+R_{\text{in}}}+R_n\right)\delta f}$$

$$= \frac{E^2R_{\text{in}}^2}{4kT(R_S+R_{\text{in}})(R_SR_{\text{in}}+R_nR_S+R_nR_{\text{in}})\,\delta f}.$$

Consider the case of an amplifier of bandwidth 10 kHz and input resistance 100 kΩ being matched to a source of 100 μV. If the equivalent noise resistance referred to the input is 40 kΩ, the equivalent noise circuit is as shown in Fig. 8.15. The signal is halved to give 50 μV at the input while the noise generated is due to an effective resistance of 40 kΩ in series with two 100 kΩ resistors in parallel, i.e. 90 kΩ.

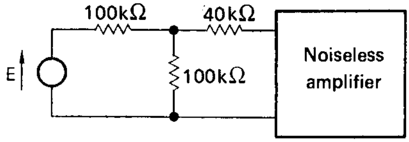


FIG. 8.15. Numerical example of Fig. 8.14.

Hence the S/N ratio at the output is given by

$$\text{S/N ratio} = \frac{(50\times 10^{-6})^2}{4\times 1{\cdot}38\times 10^{-23}\times 290\times 90\times 10^3\times 10^4}$$

$$= 17{\cdot}5 \quad \text{or} \quad 22{\cdot}4 \text{ dB}.$$

The value of $1{\cdot}38\times 10^{-23}$ J/K is assumed for Boltzmann's constant, and the absolute temperature is taken to be 17°C or 290 K.

Although the circuit shown in Fig. 8.14 appears identical to that of a transistor amplifier with resistors R_n and R_{in} replacing R_{n_1} and R_{n_2} respectively, it must be remembered that R_{n_2} represents a current generator and does not act as a potential divider with the source resistance R_S. This is best seen by the noise equivalent circuit shown in Fig. 8.16.

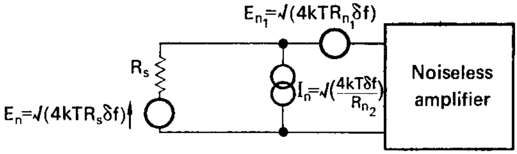

$E_{n_1}=\surd(4kTR_{n_1}\delta f)$
R_s
$E_n=\surd(4kTR_s\delta f)$
$I_n=\surd\left(\frac{4kT\delta f}{R_{n_2}}\right)$
Noiseless amplifier

FIG. 8.16. S/N ratio of amplifier with voltage and current noise generators.

The total noise voltage referred to the input is then given by

$$V_n^2 = E_n^2 + E_{n_1}^2 + I_n^2R_S^2 + 2\gamma E_{n_1}I_nR_S\,.$$

The term $2\gamma E_{n_1}I_nR_S$ is due to the correlation between the noise sources E_{n_1} and I_n, i.e. not being completely random and giving an additional cumulative effect,

i.e.
$$V_n^2 = 4kTR_S\,\delta f + 4kTR_{n_1}\,\delta f + \frac{4kT\,\delta f}{R_{n_2}}R_S^2$$
$$+2\gamma\,4kTR_S\,\delta f\sqrt{\left(\frac{R_{n_1}}{R_{n_2}}\right)}.$$

Hence, since the signal voltage at the input is not affected by the noise generators,

$$\text{S/N ratio} = \frac{E^2}{4kT\left(R_S + R_{n_1} + \frac{R_S^2}{R_{n_2}} + 2\gamma R_S\sqrt{\left(\frac{R_{n_1}}{R_{n_2}}\right)}\right)\delta f}.$$

8.8. Noise Factor

As stated previously, the noise generated within an amplifier is an important factor in its performance. Some means of comparing amplifiers is needed, and the usual methods employed are noise factor (or noise figure) and noise temperature. This section discusses noise factor while section 8.9 considers noise temperature.

Two definitions of noise factor are in current use:

i.e.
$$\text{noise factor} = \frac{\text{output S/N ratio of ideal amplifier}}{\text{output S/N ratio of actual amplifier}}$$

or
$$\text{noise factor} = \frac{\text{S/N ratio at input}}{\text{S/N ratio at output}}.$$

Since the S/N ratio at the input corresponds to the S/N ratio at the output of an ideal amplifier, these definitions obviously lead to the same result. The term noise figure is often used as an alternative to noise factor, but the strict definition is

$$\text{noise figure} = 10\log_{10}(\text{noise factor}).$$

The question that arises from the first definition is: What is meant by an ideal amplifier? This is defined as an amplifier where the only source of noise is the thermal noise developed in the source resistance. The S/N ratio of such an amplifier is, as shown in the previous section, given by

$$\text{S/N ratio} = \frac{E^2}{4kTR_S\,\delta f}.$$

Three other cases were also considered in section 8.7: an amplifier where the only source of noise was the input resistance, an amplifier with a finite input resistance and noise resistance, and a transistor amplifier with series-voltage and parallel-current noise generators.

In the first of these cases the S/N ratio was $\dfrac{E^2R_{\text{in}}}{4kT(R_S+R_{\text{in}})R_S\,\delta f}$, and the noise factor (NF) is then given by

$$\text{NF} = \frac{E^2/(4kTR_S\,\delta f)}{E^2R_{\text{in}}/[4kT(R_S+R_{\text{in}})R_S\,\delta f]}$$

$$= 1+R_S/R_{\text{in}}.$$

This shows that the noise factor of a correctly matched receiver (i.e. $R_{\text{in}} = R_S$) has a best possible value of 2, i.e. a noise figure of 3 dB.

The second case gave a S/N ratio of

$$\frac{E^2R_{\text{in}}^2}{4kT(R_S+R_{\text{in}})(R_SR_{\text{in}}+R_nR_S+R_nR_{\text{in}})\,\delta f},$$

and the noise factor is then given by

$$\text{NF} = \frac{E^2/(4kTR_S\,\delta f)}{(E^2R_{\text{in}}^2)/[4kT(R_S+R_{\text{in}})(R_SR_{\text{in}}+R_nR_S+R_nR_{\text{in}})\,\delta f]}$$

$$= \left(1+\frac{R_S}{R_{\text{in}}}\right)\left(1+\frac{R_n}{R_S}+\frac{R_n}{R_{\text{in}}}\right).$$

If the amplifier is matched (i.e. $R_{\text{in}} = R_S$), this expression reduces to $2+4(R_n/R_S)$.

In the case of the amplifier with series-voltage and parallel-current noise generators, the S/N ratio was shown to be given by $\dfrac{E^2}{4kT\left[R_S+R_{n_1}+(R_S^2/R_{n_2})+2\gamma R_S\sqrt{\left(\dfrac{R_{n_1}}{R_{n_2}}\right)}\right]\delta f}$. The noise factor of such an amplifier is given by

$$\mathrm{NF} = \frac{E^2/(4kTR_S\,\delta f)}{E^2/4kT\left[R_S+R_{n_1}+(R_S^2/R_{n_2})+2\gamma R_S\sqrt{\left(\dfrac{R_{n_1}}{R_{n_2}}\right)}\right]\delta f}$$

$$= 1+\frac{R_{n_1}}{R_S}+\frac{R_S}{R_{n_2}}+2\gamma\sqrt{\frac{R_{n_1}}{R_{n_2}}}.$$

The noise factor obviously depends on R_S, and its optimum value may be found by differentiating and equating to zero,

i.e. $$\frac{d(\mathrm{NF})}{dR_S} = -\frac{R_{n_1}}{R_S^2}+\frac{1}{R_{n_2}} = 0$$

or $$R_{S_{\mathrm{opt}}} = \sqrt{(R_{n_1}R_{n_2})}.$$

With this value of source resistance, the minimum noise factor is given by

$$\mathrm{NF}_{\mathrm{min}} = 1+\sqrt{\left(\frac{R_{n_1}}{R_{n_2}}\right)}+\sqrt{\left(\frac{R_{n_1}}{R_{n_2}}\right)}+2\gamma\sqrt{\left(\frac{R_{n_1}}{R_{n_2}}\right)}$$

$$= 1+2(1+\gamma)\sqrt{\left(\frac{R_{n_1}}{R_{n_2}}\right)}.$$

In the case when $\gamma = 0$,

$$\mathrm{NF} = 1+2\sqrt{\frac{R_{n_1}}{R_{n_2}}}.$$

Also $$R_{n_1} = \tfrac{1}{2}(\mathrm{NF}_{\mathrm{min}}-1)R_{S_{\mathrm{opt}}}$$

and $$R_{n_2} = 2R_{S_{\mathrm{opt}}}/(\mathrm{NF}_{\mathrm{min}}-1).$$

Hence, in general, the noise factor of such an amplifier is given by

$$\begin{aligned}\mathrm{NF} &= 1+\frac{1}{2}(\mathrm{NF_{min}}-1)R_{S_{\mathrm{opt}}}/R_S+\frac{1}{2}(\mathrm{NF_{min}}-1)R_S/R_{S_{\mathrm{opt}}}\\ &= 1+\frac{1}{2}(\mathrm{NF_{min}}-1)\left(\frac{R_{S_{\mathrm{opt}}}}{R_S}+\frac{R_S}{R_{S_{\mathrm{opt}}}}\right).\end{aligned}$$

A similar approach may be used with current amplifiers. The ideal current amplifier is one with zero input impedance and whose only source of noise is the thermal noise associated with the source conductance.

Then

$$\begin{aligned}\text{S/N ratio} &= I^2/I_n^2\\ &= I^2/(4kTG_S\,\delta f).\end{aligned}$$

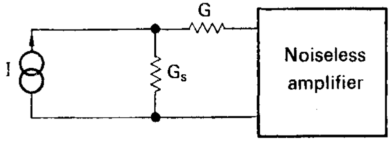


FIG. 8.17. S/N ratio of current amplifier.

With a series conductance G as shown in Fig. 8.17, the signal current is reduced in the ratio $\dfrac{G}{G+G_S}$ and the effective noise conductance becomes $\left(\dfrac{GG_S}{G+G_S}\right)$.

Hence the S/N ratio $= \dfrac{[G/(G+G_S)]^2 I^2}{4kT[(GG_S)/(G+G_S)]\,\delta f}$.

Hence the noise factor is given by

$$\begin{aligned}\mathrm{NF} &= \frac{I^2/(4kTG_S\,\delta f)}{[G(G+G_S)^2]\,I^2/\{4kT[(GG_S)/(G+G_S)]\,\delta f\}}\\ &= 1+\frac{G_S}{G}.\end{aligned}$$

This expression is the dual of that obtained for a voltage amplifier with a source resistance R_S and input resistance R_{in}. Duals of the other cases may be drawn and the noise factor estimated in each case.

8.8.1. Noise factor of a complete system

The case often arises when several networks, amplifiers, or other items of electronic equipment, of known noise factors, are connected in cascade. The noise factor of the complete system may be derived as follows.

Consider a network N_1 with a noise factor F_1 and gain G_1. If W_1 is the noise output power for an input noise power W_0,

$$F_1 = \frac{W_1}{G_1 W_0} \quad \text{or} \quad W_1 = F_1 G_1 W_0.$$

$$\begin{aligned} \text{Hence the internal noise generated} &= W_1 - G_1 W_0 \\ &= G_1 W_0 (F_1 - 1). \end{aligned}$$

Similarly, for a network N_2 with a noise factor F_2 and gain G_2, if W_2 is the noise output power for an input noise power W_0

$$F_2 = \frac{W_2}{G_2 W_0} \quad \text{or} \quad W_2 = F_2 G_2 W_0,$$

$$\begin{aligned} \text{and internal noise generated} &= W_2 - G_2 W_0 \\ &= G_2 W_0 (F_2 - 1). \end{aligned}$$

When the networks are connected in cascade as shown in Fig. 8.18, an input of W_0 gives rise to a noise power of $W_1 = F_1 G_1 W_0$ at the output of N_1 and $G_2 W_1 = F_1 G_1 G_2 W_0$ plus internal noise $G_2 W_0 (F_2 - 1)$ at the output of N_2,

i.e. noise output power $= F_1 G_1 G_2 W_0 + G_2 W_0 (F_2 - 1)$.

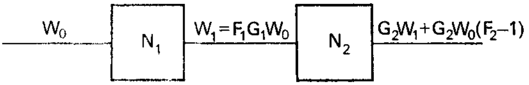


FIG. 8.18. Noise factor of networks in cascade.

If the networks were noise-free, the output noise power would be $G_1G_2W_0$.

Hence

$$\text{overall noise factor} = \frac{F_1G_1G_2W_0+G_2W_0(F_2-1)}{G_1G_2W_0}$$

$$= F_1+\frac{F_2-1}{G_1}.$$

This expression indicates the desirability of using a high-gain low-noise device as the first stage of a multi-stage amplifier.

8.9. Measurement of Noise Factor

The use of a saturated diode for the measurement of noise factor has already been mentioned in section 8.2. The actual circuit used is illustrated by Fig. 8.19 and the resistor R_S is chosen to be the same as that normally associated with the amplifier.

The first step in the measurement is to determine the output noise power P_n when there is no heater current flowing. This will

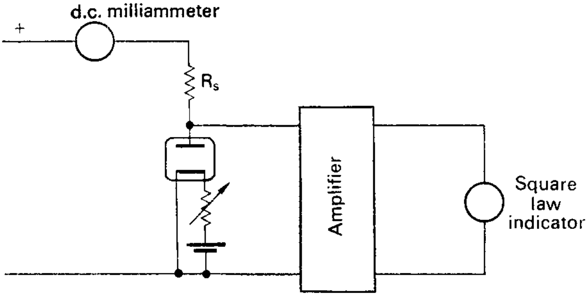


FIG. 8.19. Measurement of noise factor.

be due to the input thermal noise $\left(\text{given by } E_{n_T}^2 = 4kT \times \left(\frac{R_S R_{\text{in}}}{R_S + R_{\text{in}}}\right) \delta f\right)$ and the internal noise.

If the heater current is now increased so that the output noise power is doubled, the increase in output power (i.e. P_n) will be due to shot noise from the diode flowing through R_S and R_{in} in parallel, where R_{in} is the input resistance. This is given by $E_{n_S}^2 = 2Ie\,\delta f\left(\frac{R_S R_{\text{in}}}{R_S + R_{\text{in}}}\right)^2$.

Hence $4kT\left(\frac{R_S R_{\text{in}}}{R_S + R_{\text{in}}}\right)\delta f +$ (internal noise referred to input) $=$ $2Ie\,\delta f\left(\frac{R_S R_{\text{in}}}{R_S + R_{\text{in}}}\right)^2$.

But for a source e.m.f. E the signal at the input $= ER_{\text{in}}/(R_S + R_{\text{in}})$,

i.e. $$\text{S/N ratio} = \frac{[ER_{\text{in}}/(R_S + R_{\text{in}})]^2}{2Ie\delta f[R_S R_{\text{in}}/(R_S + R_{\text{in}})]^2} = \frac{E^2}{2Ie\,\delta f R_S^2}.$$

If the amplifier were ideal, the S/N ratio is given by $E^2/(4kTR_S\,\delta f)$.

Hence $$\text{NF} = \frac{E^2/(4kTR_S\,\delta f)}{E^2/(2Ie\,\delta f R_S^2)} = \frac{IeR_S}{2kT}.$$

Substituting for e, k and assuming the absolute temperature is 290 K,

$$\text{NF} = 20IR_S.$$

This expression shows that if R_S is known, the d.c. milliammeter may be calibrated directly in noise factor.

Alternatively, a signal generator may be connected to the input to exactly double the output power. If the output of the signal generated under these conditions is V,

$$\text{NF} = \frac{V^2}{4kTR_S\,\delta f}.$$

The latter method, however, involves the accurate measurement of the small voltage V (a few microvolts) and the noise bandwidth.

8.10. Excess Noise Temperature

Excess noise temperatures are sometimes used as an alternative to noise factors, particularly by physicists. Assuming an amplifier has a resistor R connected across its input terminals at a temperature of absolute zero, the noise output power will be entirely due to the noise generated within the amplifier. If the temperature of the resistor is now increased to temperature T_n such that the output power is doubled, the additional noise output power is due to the resistor and may be equated to the internal noise.

Then the internal noise power referred to the input $= kT_n\,\delta f$. T_n is known as the excess noise temperature of the amplifier.

At any absolute temperature T the total noise referred to the input is given by

$$P_n = kT\,\delta f + kT_n\,\delta f = k(T+T_n)\,\delta f.$$

Then noise factor,
$$F = \frac{k(T+T_n)\,\delta f}{kT\,\delta f}$$
$$= \frac{T+T_n}{T}$$

or
$$T_n = (F-1)T.$$

This expression gives the relation between excess noise temperature and noise factor. The relation between the two may be extended to the case when a composite noise factor is considered,

i.e.
$$F = F_1 + \frac{F_2-1}{G_1}.$$

But $$F = \frac{T+T_n}{T}, \quad F_1 = \frac{T+T_{n_1}}{T}, \quad F_2 = \frac{T+T_{n_2}}{T}.$$

Therefore $$\frac{T+T_n}{T} = \frac{T+T_{n_1}}{T} + \frac{\frac{T+T_{n_2}}{T}-1}{G_1},$$

i.e. $$T_n = T_{n_1} + \frac{T_{n_2}}{G_1}.$$

This expression indicates the advantage of using a high-gain low-noise preamplifier in front of a relatively noisy amplifier. Consider the case of an amplifier with a gain of 40 dB (10,000 times as a power ratio) and an excess noise temperature of 500 K. At a temperature of 17°C and a bandwidth of 10 kHz, the total noise referred to the input is $k(T+T_n)\,\delta f$

$$= 1{\cdot}38 \times 10^{-23} \times (290+500) \times 10^4$$
$$= 1{\cdot}09 \times 10^{-16}\ \text{W}.$$

With a preamplifier having a gain of 20 dB (100 times as a power ratio) and an excess noise temperature of 100 K, the composite excess noise temperature is $100+\frac{500}{100} = 105$ K, and the total noise referred to the input is

$$1{\cdot}38 \times 10^{-23}(290+105) \times 10^4 = 0{\cdot}525 \times 10^{-16}\ \text{W}.$$

This corresponds to a reduction of approximately 3 dB in the composite noise figure.

The concept of noise temperature is not limited to amplifiers. When a resistor generates more noise than the expected thermal noise (e.g. when it carries d.c. current), the effect may be expressed in terms of an excess noise temperature,

i.e. $$T_n = \frac{E_n^2}{4kR\,\delta f} - T,$$

where E_n is the actual noise voltage generated, R is the true resistance, and T is the absolute ambient temperature.

In communication work aerials are also treated in this fashion.

Questions on Chapter 8

Q.1. Discuss the various sources of noise in valve and transistor amplifiers.

Q.2. Explain what is meant by equivalent noise resistance. A source of internal resistance 50 Ω is matched to a high gain amplifier. If the bandwidth is 40 kHz and the output S/N ratio is 36 dB, find the source e.m.f. Assume a temperature of 17°C, $k = 1{\cdot}38 \times 10^{-23}$ J/K, and ignore internal noise.

Q.3. Determine the number of stages in an amplifier if the full output of 10 V r.m.s. is developed due to noise. Each stage has a bandwidth of 10 MHz, the input resistance is 3 kΩ, and the equivalent noise resistance referred to the input is 1·5 kΩ. The temperature is 27°C, Boltzmann's constant is $1{\cdot}37 \times 10^{-23}$ J/K and the source resistance is 2 kΩ. The voltage gain per stage is ten times.

Q.4. Show that if the equivalent noise generators of an amplifier are V_n and I_n in series and parallel with the input respectively, the optimum value of source resistance is given by $R_S = V_n/I_n$.

Q.5. Determine the noise factor of an amplifier whose input resistance is 10 kΩ and noise resistance referred to the input is 2 kΩ when driven from a source of resistance 5 kΩ.

Q.6. Show that the noise factor of a current amplifier whose only source of noise is the input conductance G_{in} is given by

$$\text{NF} = 1 + \frac{G_{\text{in}}}{G_S}.$$

Q.7. Distinguish between noise factor and noise temperature and derive the relationship between them.

CHAPTER 9

Linear Circuits

LINEAR circuits for valves and bipolar transistors have been covered in detail in previous volumes. In this chapter some corresponding circuits for field effect transistors will be considered and more advanced applications discussed.

9.1. Bias Circuits for the FET

The bias circuits used for the FET follow the same lines as those for vacuum valves and these are illustrated by Fig. 9.1 (a) (b), and (c). Figure 9.1 (a) shows an *n*-channel junction device with a standing bias being obtained from an auxiliary supply. An alternative, and more usual, method is shown in Fig. 9.1 (b), where the drain current flowing through the source resistor R_S provides a bias $I_D R_S$.

Figure 9.1 (c) illustrates an *n*-channel MOST operating in the enhancement mode, the forward bias on the gate being derived via a high resistance R_G connected between drain and gate, typically of the order of 27 MΩ.

9.2. Methods of Connection for the FET

The three possible methods of connection for the FET are illustrated by Figs. 9.2, 9.3, and 9.4. Figure 9.2 shows the equivalent circuit of a common-source stage, the circuit of which is given in Fig. 9.1 (b).

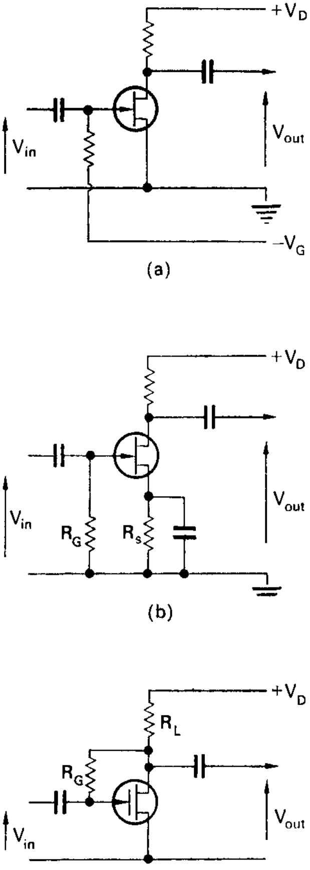


FIG. 9.1. FET bias circuits.

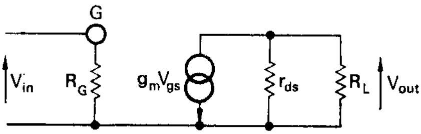


FIG. 9.2. Equivalent circuit of Fig. 9.1 (b).

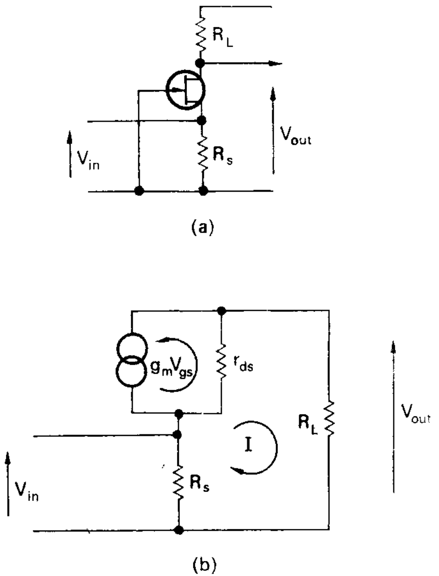


FIG. 9.3. Common-gate circuit.

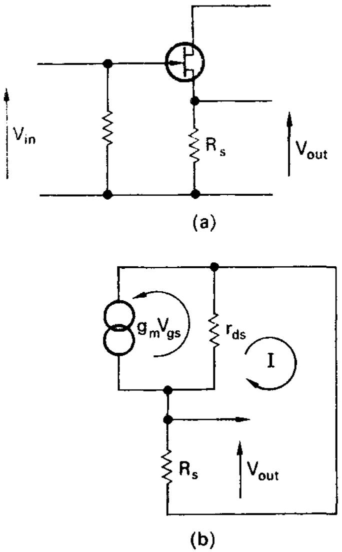


FIG. 9.4. Common-drain circuit.

Obviously, in this case, $V_{gs} = V_{\text{in}}$

and $$V_{\text{out}} = -\frac{g_m V_{gs} r_{ds} R_L}{r_{ds}+R_L},$$

i.e. $$\text{voltage gain} = \frac{V_{\text{out}}}{V_{\text{in}}} = -\frac{g_m r_{ds} R_L}{r_{ds}+R_L}$$

Figure 9.3 (a) and (b) illustrates the circuit and equivalent circuit of a FET connected in the common-gate configuration.

In this case,

$$V_{gs} = -V_{\text{in}}$$

and $$V_{\text{in}} = I(r_{ds}+R_L)+g_m V_{gs} r_{ds},$$

i.e. $$I = \frac{V_{\text{in}}(1+g_m r_{ds})}{(r_{ds}+R_L)}.$$

Also $$V_{\text{out}} = IR_L.$$

Therefore $$\text{gain} = \frac{V_{\text{in}}}{V_{\text{out}}}$$

$$= \frac{(1+g_m r_{ds})}{(r_{ds}+R_L)} R_L.$$

The third method of connection is the source follower shown with its equivalent circuit in Fig. 9.4.

$$V_{gs} = V_{\text{in}}+IR_S.$$

Also $$I(R_S+r_{ds})+g_m V_{gs} r_{ds} = 0,$$

i.e. $$I(R_S+r_{ds})+g_m r_{ds}(V_{\text{in}}+IR_S) = 0$$

or $$I = -\frac{g_m r_{ds} V_{\text{in}}}{R_S+r_{ds}+g_m r_{ds} R_S}.$$

Therefore $$\text{gain} = -\frac{IR_S}{V_{\text{in}}}$$

$$= \frac{g_m r_{ds} R_S}{r_{ds}+R_S(1+g_m r_{ds})}.$$

9.3. FET Circuits

In this section some typical FET circuits will be considered. Space precludes an exhaustive treatment, but by analogy with valve circuits it should prove sufficient to enable the reader to analyse the majority of FET circuits.

9.3.1. Direct-coupled FET amplifier

A typical FET d.c. amplifier is shown in Fig. 9.5 and its equivalent circuit in Fig. 9.6.

The equations for this circuit are:

$$I_1(R+R_S+r_{ds})+I_2R_S+g_mV_{gs_1}r_{ds} = 0, \qquad (9.3.1)$$
$$I_2(R+R_S+r_{ds})+I_1R_S+g_mV_{gs_2}r_{ds} = 0, \qquad (9.3.2)$$
$$V_{\text{out}} = I_1R-I_2R = (I_1-I_2)\,R. \qquad (9.3.3)$$

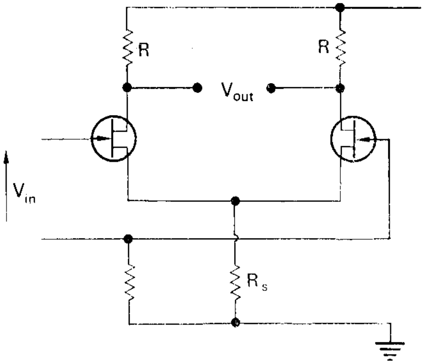


FIG. 9.5. Direct-coupled FET amplifier.

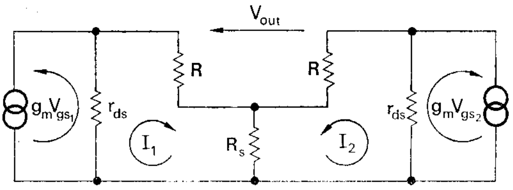


FIG. 9.6. Equivalent circuit of Fig. 9.5.

Subtracting eqn. (9.3.2) from eqn. (9.3.1),

$$I_1(R+r_{ds})-I_2(R+r_{ds})+g_m r_{ds}(V_{gs_1}-V_{gs_2}) = 0.$$

But $$(I_1-I_2) = \frac{V_{\text{out}}}{R} \quad \text{and} \quad (V_{gs_1}-V_{gs_2}) = V_{\text{in}}.$$

Therefore $$\frac{V_{\text{out}}}{R}(R+r_{ds})+g_m r_{ds} V_{\text{in}} = 0,$$

i.e. $$\text{gain} = \frac{V_{\text{out}}}{V_{\text{in}}} = -\frac{g_m r_{ds} R}{R+r_{ds}}.$$

9.3.2. *A.F. FET amplifier*

To illustrate the use of the FET in a.f. circuits consider the simple amplifier shown in Fig. 9.7.

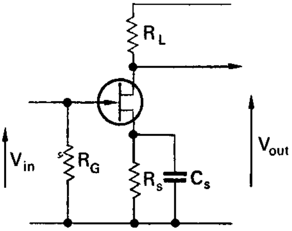


FIG. 9.7. A.F. FET amplifier.

At mid band the gain will be

$$A = -\frac{g_m R_L r_{ds}}{R_L+r_{ds}}.$$

At low frequencies the equivalent circuit is as shown in Fig. 9.8.

$$V_g = V_{\text{in}},$$

$$V_s = -I\left(\frac{R_S}{1+j\omega C_S R_S}\right),$$

$$V_{gs} = V_g-V_s = V_{\text{in}}+I\left(\frac{R_S}{1+j\omega C_S R_S}\right).$$

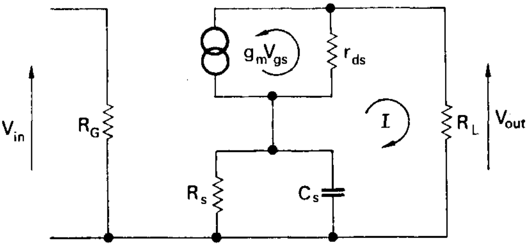


FIG. 9.8. Equivalent circuit of Fig. 9.7.

Therefore

$$I\left[R_L+r_{ds}+\frac{R_S}{1+j\omega C_S R_S}\right]+g_m r_{ds}\left[V_{\text{in}}+\frac{IR_S}{1+j\omega C_S R_S}\right]=0.$$

Therefore $I=-\dfrac{g_m r_{ds} V_{\text{in}}}{R_L+r_{ds}+\left(\dfrac{R_S}{1+j\omega C_S R_S}\right)+g_m r_{ds}\left(\dfrac{R_S}{1+j\omega C_S R_S}\right)}.$

Therefore

$$\frac{V_{\text{out}}}{V_{\text{in}}}=-\frac{g_m r_{ds} R_L}{R_L+r_{ds}+\left(\dfrac{R_S}{1+j\omega C_S R_S}\right)+g_m r_{ds}\left(\dfrac{R_S}{1+j\omega C_S R_S}\right)},$$

i.e. gain $=\dfrac{A}{1+\left(\dfrac{R_S}{1+j\omega C_S R_S}\right)\left(\dfrac{1+g_m r_{ds}}{R_L+r_{ds}}\right)}$

$$=\frac{A(1+j\omega C_S R_S)}{\left[\left(1+\dfrac{R_S(1+g_m r_{ds})}{R_L+r_{ds}}\right)+j\omega C_S R_S\right]},$$

i.e., at the 3 dB point,

$$\frac{\sqrt{(1+\omega^2 C_S^2 R_S^2)}}{\sqrt{\left[\left(1+\dfrac{R_S(1+g_m r_{ds})}{R_L+r_{ds}}\right)^2+\omega^2 C_S^2 R_S^2\right]}}=\frac{1}{\sqrt{2}},$$

$$2+2\omega^2 C_S^2 R_S^2=\left(1+\frac{R_S(1+g_m r_{ds})}{R_L+r_{ds}}\right)^2+\omega^2 C_S^2 R_S^2,$$

or $$\omega^2 C_S^2 R_S^2 = \left(1+\frac{R_S(1+g_m r_{ds})}{R_L+r_{ds}}\right)^2 - 2,$$

i.e. $$f = \frac{1}{2\pi C_S R_S}\sqrt{\left[\left(1+\frac{R_S(1+g_m r_{ds})}{R_L+r_{ds}}\right)^2 - 2\right]}.$$

If r_{ds} is high, $$f = \frac{g_m}{2\pi C_S}.$$

This gives the low frequency at which the gain falls to $1/\sqrt{2}$ of the medium-frequency value.

9.3.3. R.F. FET amplifier

The use of the FET in r.f. amplifier circuits is illustrated by Fig. 9.9. A grounded-gate circuit ensures separation between output (drain) and input (source) circuits.

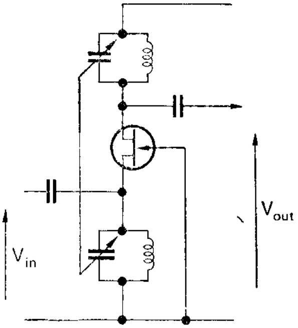


FIG. 9.9. FET common-gate r.f. amplifier.

As in the case of section 9.2 it may be shown that at resonance the gain is given by

$$A = \frac{(1+g_m r_{ds})R_D}{r_{ds}+R_D},$$

where R_D is the dynamic impedance of the tuned circuit.

If r_{ds} is high, this reduces to

$$A = g_m R_D.$$

9.3.4. FET–bipolar transistor circuits

Useful combinations of FETs and bipolar transistors are possible that combine the advantages of the high input impedance of a FET and the high current gain of a bipolar transistor. One such circuit is illustrated in Fig. 9.10, and its equivalent circuit is shown in Fig. 9.11. To a first approximation, r_{ds}, h_{re}, and $1/h_{oe}$ may be ignored, giving the equivalent circuit shown in Fig. 9.12.

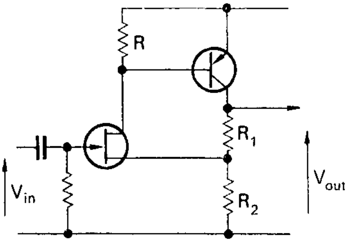


FIG. 9.10. FET–bipolar combination.

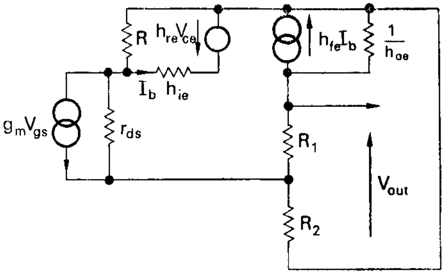


FIG. 9.11. Equivalent circuit of Fig. 9.10.

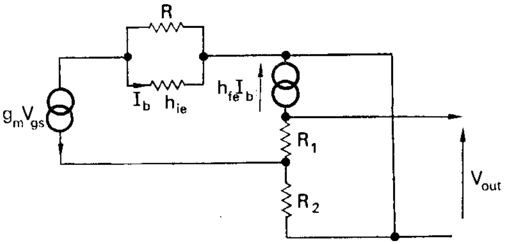


FIG. 9.12. Simplified equivalent circuit of Fig. 9.10.

From the diagram

$$I_b = -\frac{R}{R+h_{ie}} g_m V_{gs},$$

$$V_{gs} = V_{\text{in}} - (g_m V_{gs} - h_{fe} I_b)\, R_2$$

$$= V_{\text{in}} - \left(g_m V_{gs} + \frac{h_{fe} R g_m V_{gs}}{R+h_{ie}}\right) R_2$$

or
$$V_{gs} = \frac{V_{\text{in}}}{1 + g_m R_2 + \dfrac{h_{fe} R R_2 g_m}{R+h_{ie}}},$$

i.e.
$$V_{gs} \triangleq V_{\text{in}} \left(\frac{R+h_{ie}}{h_{fe} R R_2 g_m}\right).$$

Therefore
$$V_{\text{out}} = g_m V_{gs} R_2 - h_{fe} I_B (R_1 + R_2)$$

and
$$I_b = -\frac{V_{\text{in}}}{h_{fe} R_2}.$$

Hence,
$$V_{\text{out}} = g_m V_{\text{in}} \left(\frac{R+h_{ie}}{h_{fe} R R_2 g_m}\right) R_2 + h_{fe} \frac{V_{\text{in}}}{h_{fe} R_2} (R_1 + R_2)$$

or
$$\text{gain} = \frac{V_{\text{out}}}{V_{\text{in}}} = \frac{(R+h_{ie})}{h_{fe} R} + \frac{(R_1 + R_2)}{R_2}$$

$$\triangleq \left(\frac{R_1 + R_2}{R_2}\right),$$

i.e. the gain depends mainly on the values of the resistors R_1 and R_2. This result may be obtained by considering the circuit as a feedback network with heavy negative feedback,

i.e.
$$A' = \frac{A}{1+\beta A}$$

$$= \frac{1}{\beta} \quad \text{if} \quad \beta A \gg 1$$

$$= \left(\frac{R_1 + R_2}{R_2}\right).$$

9.4. Feedback

Feedback has already been considered in detail in Volume 2 when the cases of voltage and current amplifiers with series, voltage and current feedback, and parallel, voltage and current feedback respectively were discussed. These are shown in Figs. 9.13–9.16.

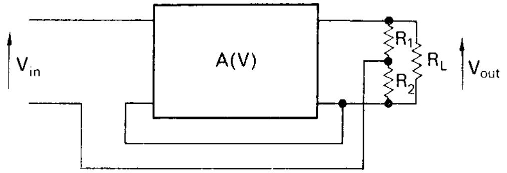


FIG. 9.13. Voltage amplifier with series-voltage feedback.

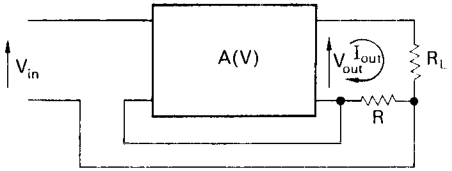


FIG. 9.14. Voltage amplifier with series-current feedback.

Figure 9.13 illustrates a voltage amplifier with series-voltage feedback, i.e. the feedback is in series with the input and proportional to the output voltage. Obviously the proportion β of the output voltage feedback is given by

$$\beta = \frac{R_2}{R_1+R_2}.$$

Figure 9.14 illustrates the case of a voltage amplifier with series-current feedback, where a voltage proportional to the output current is fed back in series with the input,

i.e. voltage fed back $= I_{\text{out}}R$

or
$$\beta V_{\text{out}} = \frac{V_{\text{out}}}{R+R_L} R,$$

i.e.
$$\beta = \frac{R}{R+R_L}.$$

One difference between this and the previous case is that with a short-circuit load there is no feedback in the case of voltage feedback, but in the case of current feedback the voltage fed back is still present.

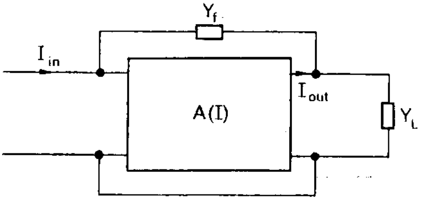


FIG. 9.15. Current amplifier with parallel-current feedback.

Figure 9.15 shows a current amplifier with parallel-current feedback, i.e. the current is fed back in parallel with the input. If the input admittance is much greater than the feedback admittance Y_f, the proportion β of the current fed back is given by

$$\beta = \frac{Y_f}{Y_f + Y_L}.$$

The final case considered is shown in Fig. 9.16. In this case the current fed back is proportional to the output voltage, i.e. current

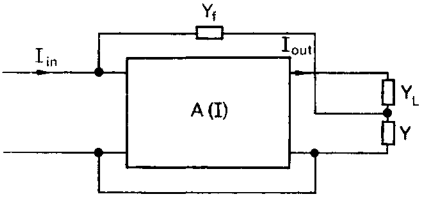


FIG. 9.16. Current amplifier with parallel-voltage feedback.

fed back

$$= V_{\text{out}} \frac{1/(Y+Y_f)}{[1/(Y+Y_f)]+(1/(Y_L)} Y_f,$$

i.e.

$$\beta I_{\text{out}} = I_{\text{out}} \frac{Y_f}{Y+Y_f}.$$

Therefore

$$\beta = \frac{Y_f}{Y+Y_f}.$$

With an open-circuit load there is no feedback in the case of voltage feedback, but for current feedback a feedback current is still present.

9.4.1. Negative feedback

As mentioned in the previous volume, negative feedback was probably the most important development in the design of stable amplifier circuits. Its effect is summarized in the Table 4.

FETs are voltage amplifiers, and typical circuits employing negative feedback are shown in Figs. 9.17 and 9.18, the former illustrating simple series-current feedback and the latter series-voltage feedback.

In the circuit shown in Fig. 9.17, the drain current flowing through the source resistance R_S provides the feedback and the gain is reduced from $-\frac{g_m r_{ds} R_L}{r_{ds}+R_L}$ to a value A',

where

$$A' = \frac{-(g_m r_{ds} R_L)/(r_{ds}+R_L)}{1+(R_S/R_L)(g_m r_{ds} R_L)/(r_{ds}+R_L)}$$

$$= -\frac{g_m r_{ds} R_L}{r_{ds}+R_L+g_m r_{ds} R_S}.$$

Figure 9.18 is an alternative version of the source follower shown in Fig. 9.4. The additional parallel *RC* combination in the source lead enables correct bias conditions to be determined for the FET.

TABLE 4

	Voltage amplifier		Current amplifier	
	Series voltage FB	Series current FB	Parallel current FB	Parallel voltage FB
Gain	reduced but stabilized			
Input impedance	increased		decreased	
Output impedance	decreased	increased	decreased	increased
Phase response	improved			
Frequency response	improved			
Distortion	reduced			

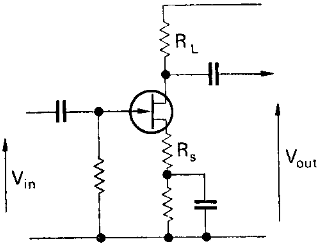


FIG. 9.17. FET amplifier with series-current feedback.

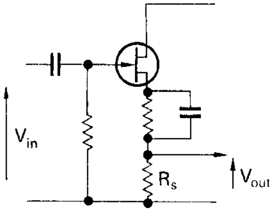


FIG. 9.18. FET amplifier with series-voltage feedback.

The gain, ignoring feedback, i.e. the input between gate and source, is

$$\frac{g_m r_{ds} R_S}{r_{ds}+R_S},$$

and all the output voltage is fed back in antiphase to the input.

Hence the voltage gain with feedback is given by

$$A' = \frac{(g_m r_{ds} R_S)/(r_{ds}+R_S)}{1+g_m r_{ds} R_S/(r_{ds}+R_S)}$$

$$= \frac{g_m r_{ds} R_S}{r_{ds}+R_S(1+g_m r_{ds})}.$$

Current and voltage negative feedback have opposite effects on the output impedance of an amplifier and often to obtain a given value of output impedance, a combination of voltage and current feedback is employed. This is illustrated for a voltage amplifier in Fig. 9.19.

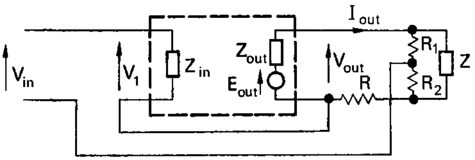


FIG. 9.19. Voltage amplifier with combined current and voltage feedback.

From Fig. 9.19,

$$V_1 = V_{in}+I_{out}R+\beta V_{out},$$

$$\left(\text{where} \quad \beta \mathrel{\hat{=}} \frac{R_1}{(R_1+R_2)}\right).$$

Therefore $$V_1 = V_{in}+V_{out}\frac{R}{Z}+\beta V_{out}.$$

But $$V_{out} = E_{out}\frac{Z}{Z_{out}+Z} \qquad (\text{for } R_1+R_2 \gg Z)$$

$$= \mu V_1 \frac{Z}{Z_{out}+Z},$$

where $$\frac{E_{out}}{V_1} = \mu.$$

Therefore $$V_{\text{out}}\left(\frac{Z_{\text{out}}+Z}{\mu Z}\right) = V_{\text{in}}+\frac{V_{\text{out}}R}{Z}+\beta V_{\text{out}}$$

or $$\text{gain} = \frac{V_{\text{out}}}{V_{\text{in}}} = \frac{1}{\left(\frac{Z_{\text{out}}+Z}{\mu Z}\right)-\frac{R}{Z}-\beta}$$

$$= \frac{\mu Z}{Z_{\text{out}}+Z-\mu R-\mu Z\beta}$$

$$= -\frac{Z}{R+\beta Z}$$

if μ is large.

If the source is replaced by its impedance Z_S and a voltage V applied to the output, as shown in Fig. 9.20, the output impedance may be evaluated.

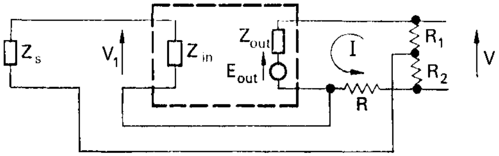


FIG. 9.20. Output impedance of a voltage amplifier with combined current and voltage feedback.

If, as is normally valid, Z_{in} is much greater than Z_S,

$$V_1 = \beta V - IR,$$

where $$\beta = \frac{R_2}{R_1+R_2}.$$

Therefore $$I = \frac{V-\mu(\beta V-IR)}{Z_{\text{out}}+R}.$$

Therefore with feedback, $$Z'_{\text{out}} = \frac{V}{I} = \frac{Z_{\text{out}}+R(1-\mu)}{1-\mu\beta}$$

$$= \frac{R}{\beta}$$

if μ is large.

Consider a three-stage voltage amplifier with a gain of 100 dB without feedback. If the optimum load is 5 kΩ and a gain of 60 dB is required with feedback present,

then
$$\frac{R}{\beta} = 5000$$

and
$$-1000 = -\frac{Z}{R+\beta Z}$$

since 60 dB corresponds to 1000 times and three stages give phase inversion.

Hence,
$$1000 = \frac{1}{(R/5000)+\beta}.$$

Therefore
$$\frac{R}{5}+1000\beta = 1,$$

i.e.
$$1000\beta+1000\beta = 1,$$

i.e.
$$\beta = \frac{1}{2000}$$

and
$$R = 2{\cdot}5\ \Omega.$$

A similar approach may be used with current amplifiers. Consider the circuit shown in Fig. 9.21. Current feedback feeds a current $\frac{Y_f}{Y_f+Y_L} I_{\text{out}}$ back to the input, while voltage feedback gives rise to a current $\frac{Y_{f_1}}{Y+Y_{f_1}} I_{\text{out}}$ flowing in the input circuit.

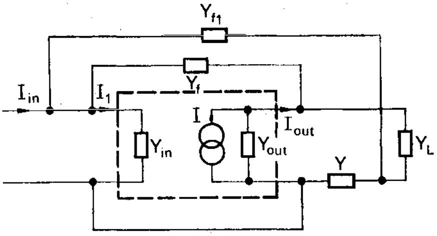


FIG. 9.21. Current amplifier with combined current and voltage feedback.

Hence $$I_1 = I_{in} + \beta I_{out} + \gamma I_{out},$$

where $$\beta = \frac{Y_f}{Y_f + Y_L} \quad \text{and} \quad \gamma = \frac{Y_{f_1}}{Y + Y_{f_1}}.$$

If $$\frac{I}{I_{in}} = \mu_I, \quad \frac{I_{out}}{I_{in}} = \frac{\mu_I Y_L}{Y_L + Y_{out}} \quad \text{very nearly.}$$

Therefore $$\text{current gain} = \frac{I_{out}}{I_{in}}$$

$$= \frac{1}{\left(\frac{Y_{out} + Y_L}{\mu_I Y_L}\right) - \beta - \gamma}$$

$$= -\frac{1}{\beta + \gamma}$$

if μ_I is high.

Since Y_{in} is assumed to very be high, the output admittance may be evaluated from the circuit shown in Fig. 9.22. In this circuit a voltage V applied to the output gives rise to a current I.

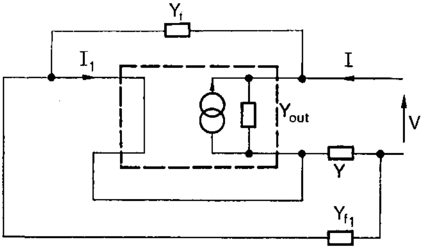


Fig. 9.22. Output admittance of a current amplifier with combined current and voltage feedback.

The current I is much greater than any current in the input circuit and the voltage across Y_f is very nearly I/Y. Hence the input current I is given by

$$I_1 = VY_f - \frac{I}{Y} Y_{f_1} \mathrel{\hat{=}} VY_f - I\gamma.$$

Then the output voltage is very nearly

$$V = \frac{(\mu_I I_1 + I)}{Y_{\text{out}}},$$

where μ_I is the short-circuit current gain,

i.e. $$VY_{\text{out}} = \mu_I(VY_f - I\gamma) + I$$

or effective output admittance

$$Y'_{\text{out}} = \frac{I}{V} = \frac{Y_{\text{out}} - \mu_I Y_f}{1 - \mu_I \gamma}.$$

If μ_I is high, $$Y'_{\text{out}} = \frac{Y_f}{\gamma}.$$

Consider a three-stage current amplifier with an optimum output admittance of 2×10^{-3} S and a current gain without feedback of 100 dB. If the feedback components to give a working gain of 60 dB are required,

$$2\times 10^{-3} = \frac{Y_f}{\gamma},$$

$$-1000 = -\frac{1}{\beta + \gamma}$$

and $$\beta = \frac{Y_f}{Y_L} = \frac{Y_f}{2\times 10^{-3}}.$$

Assuming the load is equal to the optimum output admittance,

$$1000 = 1\bigg/\left(\frac{Y_f}{2\times 10^{-3}} + \frac{Y_f}{2\times 10^{-3}}\right),$$

i.e. $$2Y_f = \frac{2\times 10^{-3}}{10^3}$$

and $$Y_f = 10^{-6}\text{ S}.$$

Since this is resistive, $R_f = 1\ \text{M}\Omega$.

Therefore $$\beta = \frac{10^{-6}}{2\times 10^{-3}} = \frac{1}{2000}\,.$$

Therefore $$\gamma = \frac{1}{2000}\,.$$

If Y_{f_1} is 10^{-4} S, $Y = 0{\cdot}2$ S, i.e. a resistance of 5 Ω.

9.4.2. *Positive feedback*

This is the basis of oscillator circuits, and as an example of the use of the FET consider the Wien bridge oscillator shown in Fig. 9.23.

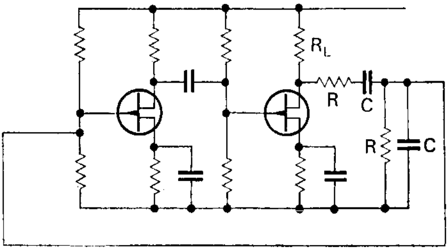


Fig. 9.23. Wien bridge FET oscillator.

The amplifier may be considered as having an internal voltage gain A between input and output; a very high input impedance compared with the frequency determining components R and X_c; and a low output impedance, very nearly R_L. The circuit may then be represented by Fig. 9.24.

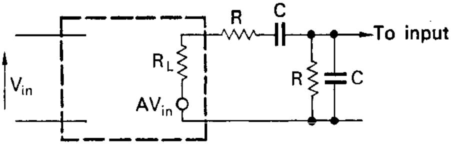


Fig. 9.24. Equivalent circuit of Fig. 9.23.

The proportion of the output e.m.f. fed back is β, where

$$\beta = \frac{R/(1+j\omega CR)}{R_L+\dfrac{(1+j\omega CR)}{j\omega C}+\dfrac{R}{(1+j\omega CR)}}.$$

For oscillation to be maintained, $\beta A = 1$,

i.e. $$\frac{j\omega CRA}{j\omega CR_L(1+j\omega CR)+(1+j\omega CR)^2+j\omega CR} = 1,$$

i.e. $$\frac{j\omega CRA}{(1-\omega^2C^2R^2-\omega^2C^2RR_L)+3j\omega CR+j\omega CR_L} = 1.$$

For no j term,

$$1-\omega^2C^2R^2-\omega^2C^2RR_L = 0,$$

i.e. $$\omega^2 = \frac{1}{C^2R^2[1+(R_L/R)]}.$$

Then $$\frac{A}{3+R_L/R} = 1,$$

i.e. $$A = 3+\frac{R_L}{R}.$$

These expressions indicate that the circuit will oscillate at a frequency $1/[2\pi CR\sqrt{(1+R_L/R)}]$ if the internal voltage gain is greater than $3+R_L/R$. In practice the voltage gain is usually much higher than this, and has to be reduced. A convenient method is to apply negative feedback from the second drain to an undecoupled resistor in the first source.

9.5. Cascade Circuits

This type of circuit is used with valves, bipolar and field effect transistors, and is characterized by one device acting as the load of another.

9.5.1. Valve cascade circuit

A valve cascade circuit is illustrated by Fig. 9.25. The lower valve is operated in the usual common cathode configuration, but the upper valve is a grounded grid stage, ensuring isolation between input and output circuits.

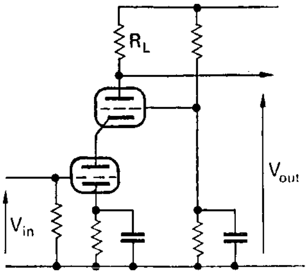


FIG. 9.25. Valve cascade circuit.

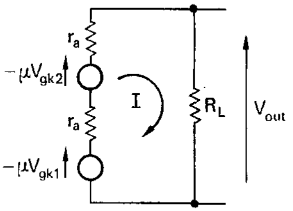


FIG. 9.26. Equivalent circuit of Fig. 9.25.

The equivalent circuit for Fig. 9.25 is shown in Fig. 9.26, where

$$
\begin{aligned}
V_{gk_1} &= V_{\text{in}}, \\
V_{gk_2} &= -(-\mu V_{gk_1} - Ir_a) \\
&= \mu V_{\text{in}} + Ir_a .
\end{aligned}
$$

Therefore
$$I = \frac{-\mu V_{\text{in}} - \mu(\mu V_{\text{in}} + Ir_a)}{2r_a + R_L}$$

or
$$I = \frac{-\mu(\mu+1)V_{\text{in}}}{r_a(\mu+2)+R_L},$$

i.e.
$$\text{gain} = \frac{V_{\text{out}}}{V_{\text{in}}} = -\frac{\mu(\mu+1)R_L}{r_a(\mu+2)+R_L}.$$

Such a circuit is found in the input circuits of some television receivers and has the advantages of very short connections between valves, i.e. a strap across the valve base and the relatively low noise level of triodes.

9.5.2. Transistor cascade circuit

Figure 9.27 gives the bipolar version of a cascade circuit. The "y" equivalent circuit of the stage is shown in Fig. 9.28. To a first approximation, y_{re} and y_{oe} may be ignored and the circuit reduces to that shown in Fig. 9.29.

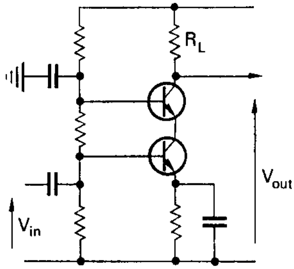


FIG. 9.27. Bipolar transistor cascade circuit.

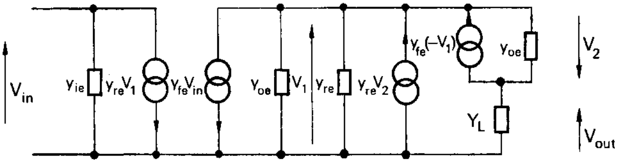


FIG. 9.28. Equivalent circuit of Fig. 9.27.

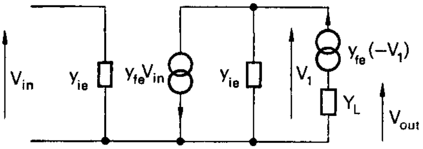


FIG. 9.29. Simplified equivalent circuit of Fig. 9.27.

Then

$$V_1 = -\frac{y_{fe}V_{\text{in}} + y_{fe}V_1}{y_{ie}}$$

or

$$V_1 = -\frac{y_{fe}V_{\text{in}}}{y_{ie} + y_{fe}}$$

and

$$\frac{V_{\text{out}}}{V_{\text{in}}} = -\frac{y_{fe}^2}{(y_{fe} + y_{ie})Y_L}.$$

9.5.3. FET cascade circuit

The circuit shown in Fig. 9.30 shows two FETs connected in cascade. This may be represented by the equivalent circuit illustrated in Fig. 9.31.

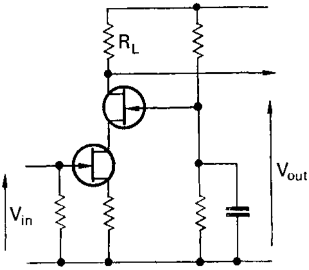


FIG. 9.30. FET cascade circuit.

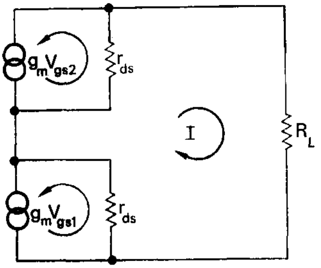


FIG. 9.31. Equivalent circuit of Fig. 9.30.

From the equivalent circuit,

$$V_{gs_2} = r_{ds}(g_mV_{\text{in}}+I).$$

Therefore $\quad I(R_L+2r_{ds})+g_mV_{\text{in}}\,r_{ds}+g_mr_{ds}^2(g_mV_{\text{in}}+I) = 0,$

i.e.
$$I = -\frac{g_mV_{\text{in}}\,r_{ds}(1+g_mr_{ds})}{R_L+r_{ds}(2+g_mr_{ds})}\,.$$

Therefore
$$\text{gain} = \frac{IR_L}{V_{\text{in}}}$$
$$= -\frac{g_mr_{ds}(1+g_mr_{ds})R_L}{R_L+r_{ds}(2+g_mr_{ds})}\,.$$

9.6. Very High Frequency and Ultra High Frequency Valve Circuits

Although semiconductor devices are becoming increasingly important at these frequencies, vacuum valves are still used extensively. This section discusses the deficiences of conventional valves and describes some circuits used. It also provides a lead-in for the VHF valves considered in Chapter 12.

9.6.1. Deficiencies of conventional valves at very high frequencies

Conventional valves suffer from three main disadvantages at very high frequencies; the Miller effect, which is covered in Volume 2 and can usually be allowed for by neutralizing circuits; the effects of cathode lead inductance, which may be reduced by using thick cathode leads; and transit time effects, which may be reduced by decreasing the cathode–anode spacing and using high anode voltages.

(a) Effects of cathode lead inductance

The effects of cathode lead inductance are best illustrated by reference to the pentode circuit shown in Fig. 9.32.

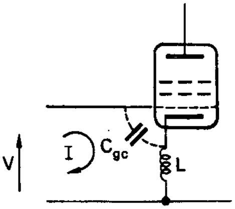


FIG. 9.32. Cathode lead inductance.

To a first approximation,

$$I_a = g_m V_{gc} = g_m \frac{I}{j\omega C_{gc}}\,.$$

Therefore $$V = \frac{I}{j\omega C_{gc}} + \left(I + \frac{g_m I}{j\omega C_{gc}}\right) j\omega L$$

or $$j\omega C_{gc} V = I[(1-\omega^2 LC_{gc}) + j\omega L g_m].$$

In practice, even though L cannot be ignored, $\omega^2 LC_{gc}$ is much less than unity.

Then the input admittance is given by

$$Y_{\text{in}} = \frac{I}{V} = \frac{j\omega C_{gc}}{1+j\omega L g_m}$$

$$= \frac{j\omega C_{gc}(1-j\omega L g_m)}{1+\omega^2 L^2 g_m^2}.$$

Also $\omega^2 L^2 g_m^2$ is much less than unity.

Therefore

$$Y_{\text{in}} = j\omega C_{gc} + \omega^2 LC_{gc} g_m.$$

Hence the input circuit may be represented by Fig. 9.33 and indicates an input resistance that is inversely proportional to the square of the frequency. This leads to damping of circuits feeding this stage.

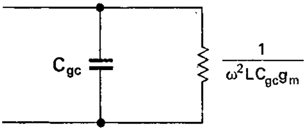


FIG. 9.33. Input circuit allowing for cathode lead inductance.

(b) Transit time effects

Consider a pentode operating at such a frequency that the transit time of electrons between cathode and anode is of the order of one-tenth of the period. If electrons take a time t_g to move between cathode and grid and a further time t_a to cover the distance between grid and anode, the instantaneous values of the currents into the grid and anode regions are given by

$$i_1 = g_m V \sin \omega(t-t_g),$$
$$i_2 = g_m V \sin \omega(t-t_g-t_a)$$

for a grid signal $V \sin \omega t$.

These are illustrated by Fig. 9.34.

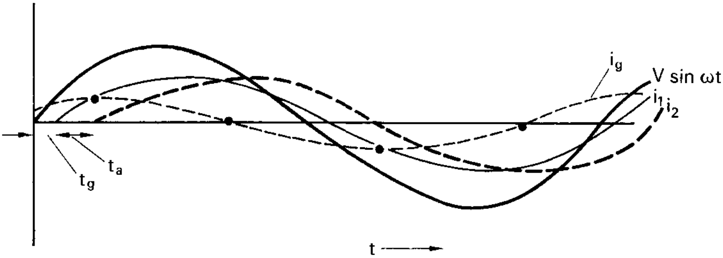


FIG. 9.34. Effect of transit time.

The difference between i_1 and i_2 is the grid current, i.e.

$$\begin{aligned} i_g &= g_m V[\sin\omega(t-t_g)-\sin\omega(t-t_g-t_a)] \\ &= 2g_m V\cos\omega(t-t_g-t_a/2)\sin\omega t_a/2 \\ &= 2g_m V\sin\frac{\omega t_a}{2}\left[\cos\omega t\cos\omega\left(t_g+\frac{t_a}{2}\right)+\sin\omega t\sin\omega\left(t_g+\frac{t_a}{2}\right)\right]. \end{aligned}$$

The in-phase component of this current is given by

$$i_p = \left[2g_m\sin\frac{\omega t_a}{2}\sin\omega\left(t_g+\frac{t_a}{2}\right)\right]V\sin\omega t$$

and corresponds to an input conductance of

$$G_{\text{in}} = 2g_m\sin\frac{\omega t_a}{2}\sin\omega\left(t_g+\frac{t_a}{2}\right).$$

If t_a and t_g are of the order of $0{\cdot}1T$, where T is the period, the sines of the angles may be replaced by the angles,

i.e. $$G_{\text{in}} = \omega^2 g_m t_a[t_g+(t_a/2)].$$

If the quadrature term is considered,

$$B_{\text{in}} = \omega g_m t_a.$$

Hence the effect of transit time is to give an input circuit as shown in Fig. 9.35. This is very similar to that due to cathode-lead inductance, and it is often difficult to distinguish between the two effects.

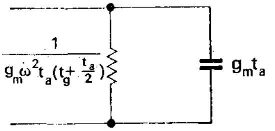


FIG. 9.35. Input circuit allowing for transit time effects.

9.6.2. *Distributed amplifiers*

Although at frequencies in the order of gigahertz, microwave valves, as discussed in Chapter 12, predominate, at somewhat lower frequencies special types of amplifier are used. One example of these is the distributed amplifier illustrated in Fig. 9.36.

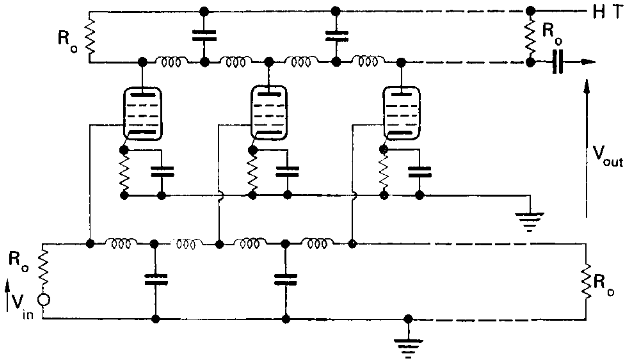


FIG. 9.36. Distributed amplifier.

Artificial delay lines are incorporated in the grid, and anode circuits and the input signal feeds each grid in sequence. The amplified anode signals split into two components that travel in opposite directions along the anode line. Half of the output power is dissipated in one terminating resistance, but signals travelling in the forward direction add up in phase to give an output voltage V_{out},

where
$$V_{\text{out}} = \frac{n}{2} I R_0,$$
$$= \frac{n}{4} g_m V_{\text{in}} R_0,$$

where n is the number of stages, I is the output current per stage, and R_0 is the characteristic impedance of the anode line.

Such a circuit provides appreciable output power with low rise times.

9.6.3. *Coaxial line circuits*

In the frequency range 300–3000 MHz, conventional inductors and capacitors become impractical, and recourse is made to transmission lines as tuning elements. As an example of such a circuit, consider the disc-seal triode used in a grounded-grid oscillator stage shown in Fig. 9.37.

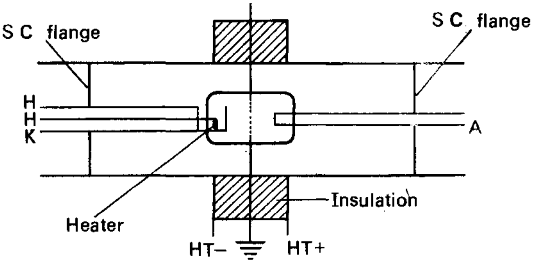


FIG. 9.37. Coaxial line oscillator.

This is basically the Colpitts oscillator illustrated in Fig. 9.38, where C_1 is the anode–cathode capacitance, L is the input impedance of the grid–anode line, and C_2 is the input impedance of the grid–cathode line. The components L and C_2 will, of course, be suitably modified by the anode–grid and grid–cathode capacitances.

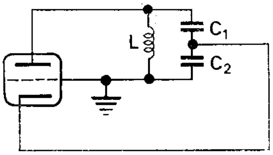


FIG. 9.38. Lumped circuit of Fig. 9.37.

Questions on Chapter 9

Q.1. Compare common-source, common-drain, and common-gate FET circuits with respect to input and output impedances and voltage gain.

Q.2. Derive an expression for the input impedance of a voltage amplifier with combined current and voltage feedback in terms of the input impedance without feedback, the open-circuit voltage gain, output and load impedance, and feedback parameters.

Q.3. Two identical triodes are connected as shown in Fig. 9.39. Derive an expression for the voltage gain of the circuit.

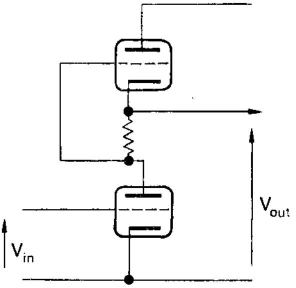


FIG. 9.39.

Q.4. The grid–cathode capacitance of a pentode operating at 100 MHz is 5 pF. If the mutual conductance is 8 mA/V and the cathode-lead inductance is 0·03 μH, determine the input admittance.

Q.5. Sketch the circuit of and determine the frequency and condition for oscillation for a Colpitts oscillator employing FETs.

CHAPTER 10

Non-linear Circuits

IN THIS chapter, the use of the field effect transistor in multivibrator circuits will be considered and sweep circuits introduced.

10.1. FET Multivibrator Circuits

The astable multivibrator, shown in Fig. 10.1, is very similar to the valve circuit discussed in detail in Chapter 12 of Volume 2. The large amount of positive feedback ensures that changes at one drain are rapidly reflected to the other gate, and fast switching occurs. The output taken from either drain is a square wave.

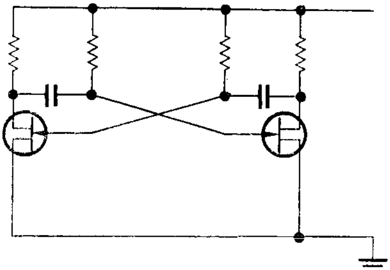


FIG. 10.1. FET astable multivibrator.

If one of the *RC* coupling networks is replaced by direct coupling, a monostable circuit results, while replacing both gives the bistable circuit familiar in counting circuits. The action of both these circuits may be seen by comparison with the corresponding valve circuits (sections 10.3 and 10.4 of Volume 2).

10.2. Characteristics of Sweep Circuits

One of the most common non-sinusoidal waveforms found in electronic circuits is the ramp function shown in Fig. 10.2. The obvious example is the time-base circuits of cathode-ray tubes.

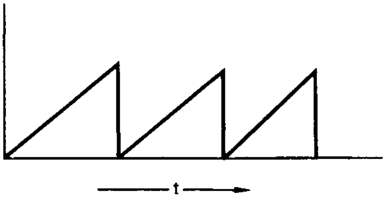


FIG. 10.2. Ideal ramp function.

The ideal waveform cannot be attained, although very close approximations may be achieved, and some means of comparing practical waveforms is desirable. Common methods used are slope error, transmission error, and displacement error—the last named being that most usually employed.

In practice, ramp functions are composed of two separate exponentials, a long time-constant curve providing the sweep and a short time-constant relation giving the very rapid flyback. This is illustrated by Fig. 10.3, which also indicates the ideal function *OA* and the sweep period *T*.

Then

$$\text{slope error} = \frac{\left(\dfrac{dV}{dt}\right)_O - \left(\dfrac{dV}{dt}\right)_A}{\left(\dfrac{dV}{dt}\right)_O}$$

$$= \frac{\tan\alpha - \tan\beta}{\tan\alpha}.$$

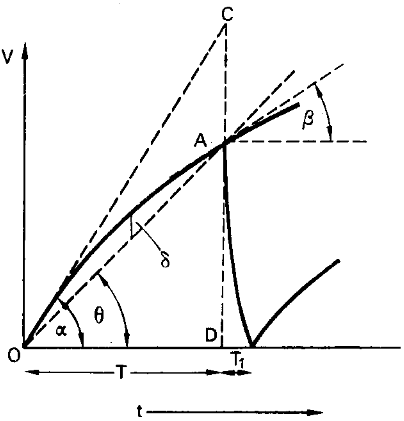


FIG. 10.3. Practical ramp function.

Displacement error or non-linearity factor

$$= \frac{\text{maximum deviation}}{\text{maximum displacement}}$$

$$= \frac{\delta}{AD}$$

$$= \frac{\delta}{T \tan \theta}.$$

Transmission error or deviation from linearity

$$= \frac{CD - AD}{CD}$$

$$= \frac{T \tan \alpha - T \tan \theta}{T \tan \alpha}$$

$$= \frac{\tan \alpha - \tan \theta}{\tan \alpha}.$$

The flyback time T_1 should be as short as possible.

10.3. Simple Ramp Function Generator

The simplest form of ramp generator is shown in Fig. 10.4. The capacitor charges exponentially until its voltage reaches the striking voltage V_S of the gas discharge tube, when it rapidly falls to the extinction voltage V_E.

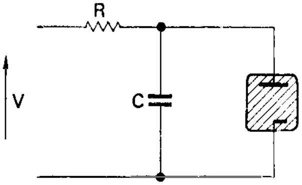


FIG. 10.4. Simple ramp function generator.

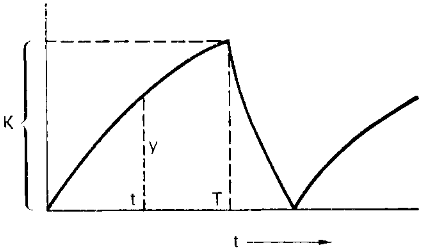


FIG. 10.5. Output waveform of Fig. 10.4.

The important part of the waveform lies between extinction and striking voltages as shown in Fig. 10.5. Then, if the amplitude is K,

$$K \propto (1-\varepsilon^{-T/RC}),$$

and at any time t

$$y \propto (1-\varepsilon^{-t/RC}).$$

Hence

$$y = \frac{K(1-\varepsilon^{-t/RC})}{(1-\varepsilon^{-T/RC})}.$$

The percentage deviation at any time t

$$= \frac{K\dfrac{(1-\varepsilon^{-t/RC})}{(1-\varepsilon^{-T/RC})}-K\dfrac{t}{T}}{K}\times 100\%$$

$$= \frac{(1-\varepsilon^{-t/RC})-\dfrac{t}{T}(1-\varepsilon^{-T/RC})}{(1-\varepsilon^{-T/RC})}\times 100\%$$

$$= \frac{\left[1-\left(1-\dfrac{t}{RC}+\dfrac{t^2}{2(RC)^2}\cdots\right)\right]-\dfrac{t}{T}\left[1-\left(1-\dfrac{T}{RC}+\dfrac{T^2}{2(RC)^2}\cdots\right)\right]}{\left[1-\left(1-\dfrac{T}{RC}+\dfrac{T^2}{2(RC)^2}\cdots\right)\right]}$$

$$\times 100\%$$

$$= \frac{\left(\dfrac{t}{RC}-\dfrac{t^2}{2(RC)^2}\cdots\right)-\left(\dfrac{t}{RC}-\dfrac{tT}{2(RC)^2}\cdots\right)}{\left(\dfrac{T}{RC}-\dfrac{T^2}{2(RC)^2}\cdots\right)}\times 100\%.$$

If high order terms are ignored,

$$\text{percentage deviation} = \frac{t}{T}\left(\frac{T-t}{2RC}\right)\times 100\%.$$

Differentiating this expression gives a maximum deviation when $t = T/2$,

i.e. $$\text{displacement error} = \frac{T}{8RC}\times 100\%.$$

10.4. The Miller Integrator

The simple RC ramp-function generator suffers from considerable non-linearity unless very high supply voltages are used. One method of improving the linearity is to increase artificially the charging time constant, and the Miller integrator is probably the best-known example of this type.

10.4.1. Miller integrator using voltage amplifier

Consider, first, the voltage amplifier form of the Miller integrator circuit shown in Fig. 10.6.

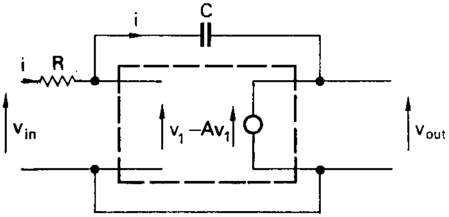


FIG. 10.6. Miller integrator incorporating voltage amplifier.

Assuming a very high input impedance

$$v_1 - v_{out} = q/C,$$

i.e.

$$-\frac{v_{out}}{A} - v_{out} = \frac{q}{C}$$

since

$$v_{out} = -Av_1.$$

If A is high,

$$q = -Cv_{out}.$$

Taking Laplace transforms,

$$\bar{q} = -C\bar{v}_{out}.$$

Also,

$$v_{in} - v_1 = iR$$

or

$$v_{in} + \frac{v_{out}}{A} = R\frac{dq}{dt}$$

since

$$i = \frac{dq}{dt}.$$

If v_{in} is a step function of amplitude V, taking Laplace transforms,

$$\frac{V}{s} + \frac{\bar{v}_{out}}{A} = R(s\bar{q} - q_0).$$

Substituting for $\bar{q}$,

$$\frac{v}{s}+\frac{\bar{v}_{\text{out}}}{A}=-RCs\bar{v}_{\text{out}}-Rq_0$$

or

$$\bar{v}_{\text{out}}=-\frac{V}{s[(1/A)+RCs]}-\frac{Rq_0}{[(1/A)+RCs]}$$

$$=-\frac{V}{CRs[s+(1/ARC)]}-\frac{Rq_0}{RC[s+(1/ARC)]}$$

$$=-\frac{V}{CR}\left(\frac{ARC}{s}-\frac{ARC}{s+(1/ARC)}\right)$$

$$-\frac{q_0}{C[s+(1/ARC)]}.$$

Taking the inverse, $v_{\text{out}}=-AV(1-\varepsilon^{-t/ARC})-\frac{q_0}{C}\varepsilon^{-t/ARC}$.

Expanding,

$$v_{\text{out}}=-AV\left[1-\left(1-\frac{t}{ARC}+\frac{t^2}{2(ACR)^2}\cdots\right)\right]$$

$$-\frac{q_0}{C}\left[1-\frac{t}{ACR}+\frac{t^2}{2(ACR)^2}\cdots\right].$$

Since $A \gg 1$,

$$v_{\text{out}}=-\frac{q_0}{C}-\frac{Vt}{CR}+\frac{Vt^2}{2A(CR)^2}\cdots.$$

The first term is a constant depending on the initial conditions, the second gives the linear change in output voltage, and the third gives the departure from linearity. The higher order terms will also contribute to the non-linearity, but their effect is minimal.

The displacement error may then be estimated by assuming that the maximum deviation occurs at the mid-point of the sweep,

i.e.

$$\text{displacement error}=\frac{[VC(T/2)^2]/[2A(CR)^2]}{VT/CR}\times 100\%$$

$$=\frac{T}{8ACR}\times 100\%.$$

Thus the displacement error is reduced by a factor A compared with the simple ramp function generator shown in Fig. 10.4.

10.4.2. Practical circuit of Miller integrator using a voltage amplifier

In a practical circuit, provision must also be made for switching the circuit at the end of the sweep, and a FET circuit is shown in Fig. 10.7. In this case, the FET is initially cut off and the capacitor is initially charged up to the supply voltage. By comparison with Fig. 10.6 it may be seen that q_0/C is equal to $-V_{DD}$.

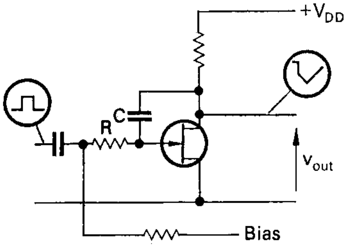


FIG. 10.7. FET version of Fig. 10.6.

A positive-going input voltage switches the device on and, after a small initial fall, the drain voltage follows the relation

$$v_{\text{out}} = V_{DD} - \frac{Vt}{CR}$$

if the second-order term is ignored.

When the input pulse falls to zero the transistor is cut off and the output voltage rises exponentially as the capacitor is again charged to the supply potential.

Pentodes may also be used for Miller circuits, but gating is normally carried out at the suppressor grid as illustrated by Fig. 10.8. In this case the control grid is taken via R to the positive end of the supply voltage and grid current flows, although the anode current is normally zero due to the negative bias applied to the suppressor grid. The capacitor C is normally charged to a few volts below the supply voltage.

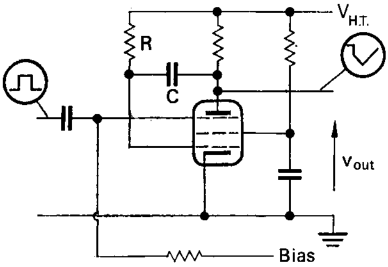


FIG. 10.8. Pentode version of Fig. 10.6.

On receipt of a positive pulse on the suppressor, anode current flows. The initial drop in anode voltage is transmitted to the control grid, which tends to maintain the anode current at a low value with the voltage drop across the anode load approximately equal to the cut-off voltage of the valve. This is virtually instantaneous and is rapidly followed by a linear increase in current and a linear decrease in anode voltage according to the law

$$V_{\text{out}} = V_{\text{H.T.}} - \frac{Vt}{CR}.$$

When the input pulse again falls, the anode current is cut off and the capacitor charges up through the anode load.

10.4.3. Miller integrator using current amplifier

The corresponding current amplifier circuit is shown in Fig. 10.9. Assuming a very low input impedance amplifier, $v_1 \to 0$,

$$q = Cv_{\text{out}}.$$

Taking Laplace transforms, $\bar{q} = C\bar{v}_{\text{out}}$.

Also
$$\begin{aligned} v_{\text{out}} &= -A_I i_1 R_L \\ &= -A_I(i_{\text{in}} + i_f)R_L. \end{aligned}$$

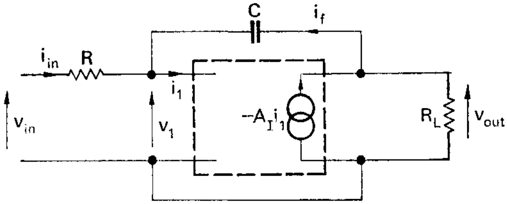


FIG. 10.9. Miller integrator incorporating current amplifier.

But $$v_{\text{in}} = i_{\text{in}}R \quad \text{and} \quad i_f = dq/dt.$$

Therefore $$v_{\text{out}} = -A_I\left(\frac{v_{\text{in}}}{R}+\frac{dq}{dt}\right)R_L.$$

Assuming v_{in} is a step function of amplitude V, Laplace transforms may be taken,

i.e. $$\frac{\bar{v}_{\text{out}}}{A_I R_L} = -\frac{V}{sR}-(s\bar{q}-q_0).$$

Substituting for $\bar{q}$,

$$\frac{\bar{v}_{\text{out}}}{A_I R_L} = -\frac{V}{sR}-sC\bar{v}_{\text{out}}+q_0$$

or $$C\bar{v}_{\text{out}}\left(s+\frac{1}{A_I R_L C}\right) = -\frac{V}{sR}+q_0,$$

$$\bar{v}_{\text{out}} = -\frac{V}{sCR[s+(1/A_I R_L C)]}+\frac{q_0}{C[s+(1/A_I R_L C)]}$$

$$= -\frac{A_I R_L V}{R}\left[\frac{1}{s}-\frac{1}{[s+(1/A_I)R_L C]}\right]+\frac{q_0}{C[s+(1/A_I R_L C)]}.$$

Taking the inverse,

$$v_{\text{out}} = -\frac{A_I R_L V}{R}[1-\varepsilon^{-t/(A_I R_L C)}]+\frac{q_0}{C}\varepsilon^{-t/(A_I R_L C)}.$$

Expanding,

$$v_{\text{out}} = -\frac{A_I R_L V}{R}\left[1-\left(1-\frac{t}{A_I R_L C}+\frac{t^2}{2(A_I R_L C)^2}\cdots\right)\right]$$
$$+\frac{q_0}{C}\left[1-\frac{t}{A_I R_L C}+\frac{t^2}{2(A_I R_L C)^2}\cdots\right].$$

Since A_I is large,

$$v_{\text{out}} = \frac{q_0}{C}-\frac{Vt}{RC}+\frac{Vt^2}{2CR(A_I R_L C)}.$$

The first term represents the initial condition, the second the linear change in output voltage, and the third the departure from linearity.

Since the maximum deviation may be assumed to occur halfway through the sweep, the displacement error may be estimated,

$$\text{i.e.}\quad \text{displacement error} = \frac{\left[V\left(\frac{T}{2}\right)^2\right]\Big/2CR(A_I R_L C)}{(VT/CR)}$$
$$= \frac{T}{8CR_L A_I}.$$

10.4.4. Practical form of Miller integrator using current amplifier

An example of this type is the bipolar transistor circuit shown in Fig. 10.10. The transistor is normally biased off and the capacitor C charged to the negative supply voltage. A negative-going pulse switches the transistor on and the output voltage rises linearly until the input voltage returns to its original value when the device is cut off. The output then falls exponentially.

The circuit shown in Fig. 10.10 does not provide a flyback that is fast enough for many purposes. To do this it is necessary to interrupt the feedback path by a circuit such as that shown in Fig. 10.11. In this circuit Q_1 normally conducts through the diode,

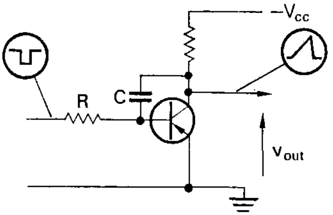


FIG. 10.10. Bipolar transistor version of Fig. 10.9.

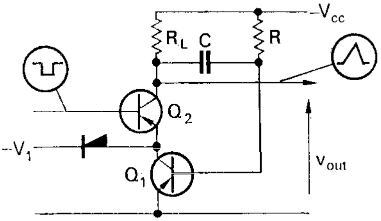


FIG. 10.11. Improved version of Fig. 10.10.

while Q_2 is cut off due to its emitter being more negative than the base. A negative pulse on the base of Q_2 switches that transistor on, and the diode is reverse-biased. The capacitor C then charges to give a linearly increasing output voltage as indicated in the previous section. When the gating pulse goes back to its original value, Q_2 is cut off and the output drops to the negative supply voltage.

10.5. Bootstrap Sweep Circuits

An alternative arrangement that provides a very nearly linear sweep is the bootstrap circuit. In this an amplifier with a voltage gain of very nearly 1 has its output fed back in series to the input as shown in Fig. 10.12.

For the input circuit

$$\begin{aligned} v_{\text{in}} &= iR + v_1 - v_{\text{out}} \\ &= R\frac{dq}{dt} - v_{\text{out}}\left(1 - \frac{1}{A}\right) \end{aligned}$$

since

$$v_{\text{out}} = Av_1.$$

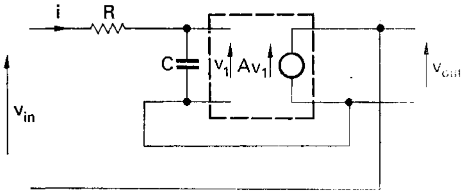


FIG. 10.12. Bootstrap sweep generator.

If the input is a positive-going step function of amplitude V, taking Laplace transforms gives

$$\frac{V}{s} = R(s\bar{q}-q_0)-\bar{v}_{\text{out}}\left(\frac{A-1}{A}\right).$$

If the capacitor is initially uncharged, $q_0 = 0$.

Also, $$v_1 = \frac{v_{\text{out}}}{A} = \frac{q}{C}.$$

Therefore taking transforms,

$$\frac{\bar{v}_{\text{out}}}{A} = \frac{\bar{q}}{C}.$$

Therefore $$\frac{V}{s} = \frac{RsC\bar{v}_{\text{out}}}{A}+\bar{v}_{\text{out}}\left(\frac{1-A}{A}\right)$$

$$= \frac{RC}{A}\bar{v}_{\text{out}}\left[s+\left(\frac{1-A}{RC}\right)\right],$$

i.e. $$\bar{v}_{\text{out}} = \frac{AV}{RC}\,\frac{1}{s\left[s+\left(\frac{1-A}{RC}\right)\right]}$$

$$= \frac{AV}{(1-A)}\left[\frac{1}{s}-\frac{1}{s+\left(\frac{1-A}{RC}\right)}\right].$$

Taking the inverse,

$$v_{\text{out}} = \frac{AV}{(1-A)}\left[1-\varepsilon^{-[(1-A)/RC]t}\right]$$

$$= \frac{AV}{(1-A)}\left\{1-\left[1-\left(\frac{1-A}{RC}\right)t+\frac{1}{2}\left(\frac{1-A}{RC}\right)^2\right]t^2\ldots\right\}$$

$$= \frac{AVt}{RC}-\frac{AV(1-A)\,t^2}{2(RC)^2}\ldots.$$

This expression shows a linearly rising function with a displacement error, again assumed half-way through the sweep.

$$\text{Displacement error} = \frac{\{[AV(1-A)]/[2(RC)^2]\}\,(T/2)^2}{(AV/RC)\,T} \times 100\%$$

$$= \frac{T(1-A)}{8RC} \times 100\%.$$

Ideally, if $A = 1$, a perfect sweep will be obtained.

10.5.1. Practical bootstrap circuit

As an example of a bootstrap sweep generator, consider the circuit shown in Fig. 10.13. Valve V_1 normally conducts heavily with a low voltage across C. A negative-going input pulse cuts the valve off and C charges the feedback from the output, raising the potential of the point X. This gives a very nearly linear charge rate of C and a very nearly linear rise in output voltage. When the input pulse returns to its original value, V_1 again conducts heavily and C rapidly discharges.

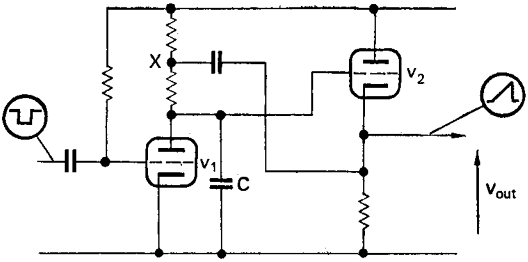


FIG. 10.13. Valve version of Fig. 10.12.

The corresponding bipolar transistor circuit is shown in Fig. 10.14. The action follows the same lines as before with the diode in the collector of Q_1 being cut off during the sweep period.

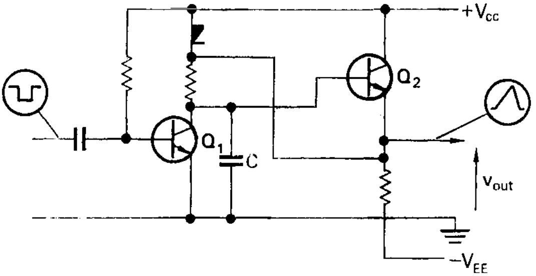


FIG. 10.14. Bipolar transistor version of Fig. 10.12.

Similar circuits are used with FETs or combinations of bipolar transistors and FETs. Figure 10.15, in fact, shows a circuit where the sweep is provided by a FET and triggering by a bipolar device.

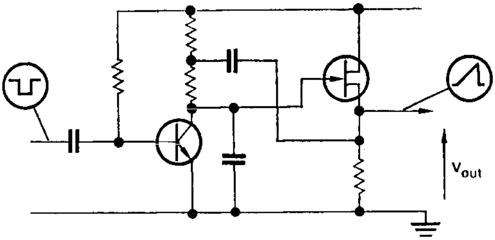


FIG. 10.15. FET bootstrap sweep generator with bipolar transistor trigger.

Questions on Chapter 10

Q.1. Discuss the use of the FET in astable, monostable, and bistable circuits.

Q.2. A particular operational amplifier has a capacitor *C* connected between output and input and a resistance *R* between input and source voltage. If the *RC* time constant is 0·05 sec and the operational amplifier has a gain of -100, estimate the maximum sweep time if the displacement error must be less than 5%.

Q.3. Derive a relationship between the voltage and current gains, integrating resistance and current amplifier load resistance of Miller integrators incorporating voltage and current amplifiers if their displacement errors are the same.

Q.4. Sketch the circuit and describe the operation of a Miller integrator incorporating bipolar transistors and external gating.

Q.5. Outline, with reference to Fig. 10.15, the action of a bootstrap sweep generator.

CHAPTER 11

Logic Circuits

ALTHOUGH logic systems may appear complex, they are in fact built up of a large number of simple, and very often identical, circuits. The similarity of these circuits and their straightforward on–off operation are important factors in the development of digital integrated circuits. In this chapter, logic operations will be considered, logic design introduced, and logic systems compared.

11.1. Logic Operations and Circuits

The basic circuits used in logic systems perform the logic operations of AND, OR, and NOT, but combinations of these circuits may be used to provide more complicated functions. The simplest combinations are circuits enabling the operations of NOT and AND (i.e. NAND) and NOT and OR (i.e. NOR) to be realized, and it is not difficult to build up circuits such as half and full adders.

11.1.1. The AND *circuit*

The AND circuit is probably most easily visualized by reference to the relay circuit shown on Fig. 11.1. With no signals on the inputs A and B, the output will be at earth potential. An input on either A or B will close one switch but the output will remain at earth potential. With both inputs energized, the relay contacts will be closed and the output will rise to a voltage, $ER_L/(R_S+R_L)$.

The actual voltage across the load R_L is relatively unimportant, and all that is really required is to know whether there is a voltage present or not. These may conveniently be represented by the levels 1 and 0, and obviously leads to a binary system.

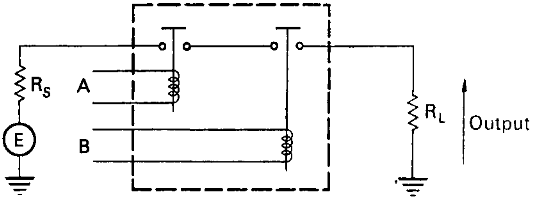


FIG. 11.1. Basic AND circuit.

In practical circuits the two levels do not, in general, correspond to 0 and 1 V. There are just two levels to be considered, and with positive logic the 1 corresponds to the higher level and the 0 to the lower level. With negative logic the situation is reversed and the 1 corresponds to the lower level.

The operation of the circuit is usually described in terms of a truth table as shown in Fig. 11.2. This shows that with no inputs on A and B (i.e. $A = 0$, $B = 0$) or with only one input (i.e. $A = 0$ $B = 1$ or $A = 1$ $B = 0$), there is no output (i.e. 0). With both inputs energized (i.e. $A = 1$, $B = 1$), there is an output of 1.

Boolean algebra may also be used to describe the operation of this circuit. For a two-input gate as shown, the operation is described symbolically by the relation

$$A \text{ and } B = A \cdot B.$$

A	B	Output
0	0	0
0	1	0
1	0	0
1	1	1

FIG. 11.2. Truth table for AND circuit.

The values of A and B are 0 or 1, but the letters are retained to denote an input without specifying whether it is 0 or 1.

It is obvious that $A \cdot B = B \cdot A$.

For a three-input gate,

$$A \text{ and } B \text{ and } C = A \cdot B \cdot C.$$

Further, if the same signal is applied to both inputs of a two-input AND gate, the output will be of the same form (i.e. $0 \cdot 0 = 0$ and $1 \cdot 1 = 1$).

Hence A and $A = A \cdot A = A$ (not A^2).

If one of the inputs of a two-input AND gate is fixed, similar relations may be stated,

i.e. $$A0 = 0 \quad \text{and} \quad A1 = A.$$

The simplest electronic form of an AND circuit is shown in Fig. 11.3. With the inputs on A, B or C all at zero potential, the diodes will carry current and the output voltage will be the minimum diode forward voltage drop. If a positive voltage (i.e. level 1) is applied to one input sufficient to cut the corresponding diode off, the output will remain in the low state due to the current flowing through the other diodes. Similarly, with positive inputs on two diodes the output will remain in the low state due to the third diode. When positive levels, which are less than the supply voltage

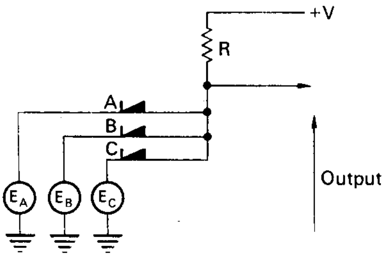


FIG. 11.3. Electronic form of AND circuit.

V are applied to all three inputs, the diodes will be conducting and the output voltage will rise to a value approaching the positive level 1. There is a slight voltage drop due to the diodes, and this limits the number of stages that may be connected in cascade.

11.1.2. The OR *circuit (inclusive* OR)

The OR circuit is also conveniently visualized by means of a relay circuit. This is shown in Fig. 11.4. With no signal on A or B, the output will be at earth potential, but an input on either A or B or both will complete the load circuit giving a positive output of $ER_L/(R_S+R_L)$.

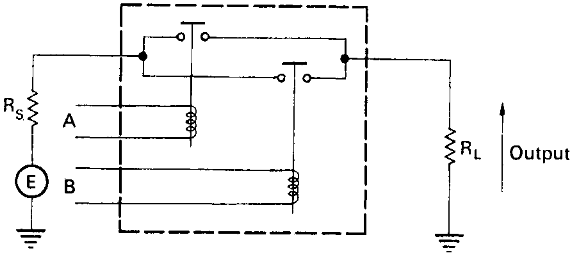


FIG. 11.4. Basic OR circuit.

A	B	Output
0	0	0
0	1	1
1	0	1
1	1	1

FIG. 11.5. Truth table for OR circuit.

As in the case of the AND circuit, it is only important to know if there is or is not a voltage present, and the corresponding truth table is as shown in Fig. 11.5. This table shows that with either or both of the inputs energized (i.e. $A = 0$ $B = 1$, or $A = 1$ $B = 0$, or $A = 1$ $B = 1$) there is an output 1.

In terms of Boolean algebra, the operation of an OR circuit may be described symbolically by the relation

$$A \text{ or } B = A+B.$$

Obviously $$A+B = B+A.$$

With a three-input gate

$$A \text{ or } B \text{ or } C = A+B+C.$$

When the same signal is applied to both inputs of a two-input OR gate, the output must be of the same form,

i.e. $$A \text{ or } A = A+A = A.$$

Again, if one of the inputs is fixed we can easily state relations for the output,

i.e. $$A+1 = 1 \quad \text{and} \quad A+0 = A.$$

A simple electronic form of OR circuit is shown in Fig. 11.6. With all the inputs zero (i.e. $A = B = C = 0$), all the diodes conduct and the output is slightly below earth potential due to the forward voltage drop of the diodes. A positive voltage on any input will give rise to a positive-going output. If this circuit is used with negative logic, it provides the AND function.

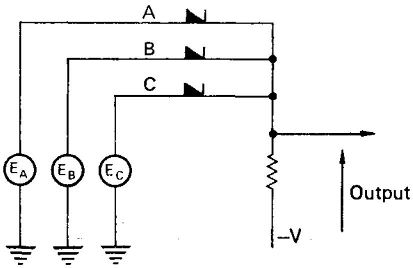


FIG. 11.6. Electronic form of OR circuit.

11.1.3. The NOT *circuit*

A circuit for performing a NOT operation is illustrated in Fig. 11.7. With no input voltage ($A = 0$) there will be an output voltage $ER_L/(R_L+R_S)$ but with the coil energized the relay contact will be opened and the output will fall to zero. Using the notation previously employed, the truth table for this circuit is shown in Fig. 11.8.

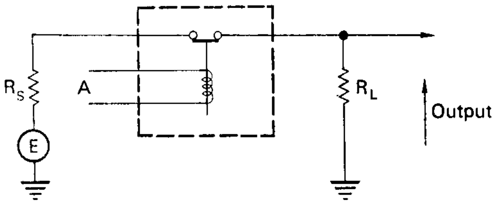


FIG. 11.7. Basic NOT circuit.

A	Output
0	1
1	0

FIG. 11.8. Truth table for NOT circuit.

The NOT operation is represented in Boolean algebra by a bar above the letter,

i.e. NOT $A = \bar{A}$.

It follows that A and $\bar{A} = A \cdot \bar{A} = 0$

and A or $\bar{A} = A+\bar{A} = 1$.

A practical form of this circuit is shown in Fig. 11.9 and is a simple inverting circuit giving an output just less than V (due to leakage current) for no input and an output close to zero for an input sufficient to bottom the transistor.

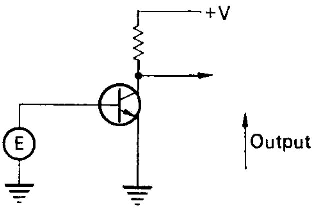


FIG. 11.9. Electronic form of NOT circuit.

11.1.4. *The* NAND *circuit*

The NAND circuit, as its name implies, is a combination of a NOT circuit and an AND circuit or, more correctly, the other way round, since the AND circuit usually comes first. A relay version of this circuit is illustrated by Fig. 11.10. With no inputs on A and B (i.e. either $A = 0$ or $B = 0$), the contacts S are closed and the output voltage is $ER_L/(R_S+R_L)$. With an input on both A and B, the contact S will be opened and the output will fall to zero.

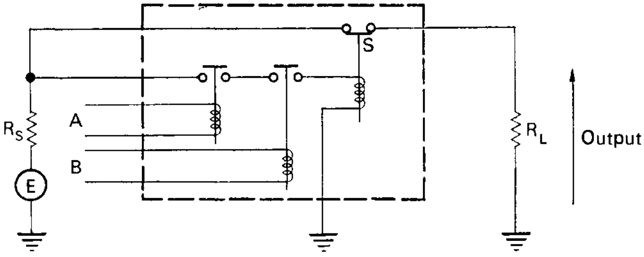


FIG. 11.10. Basic NAND circuit.

The truth table for such a circuit is shown in Fig. 11.11, and in terms of Boolean algebra the operation may be described by the relation

$$\text{NOT } (A \text{ and } B) = \text{NOT } A \text{ or NOT } B,$$

i.e.

$$\overline{A \cdot B} = \bar{A}+\bar{B}.$$

A	B	Output
0	0	1
0	1	1
1	0	1
1	1	0

FIG. 11.11. Truth table for NAND circuit.

A practical form of a NAND circuit is illustrated by Fig. 11.12. If any of the inputs are at zero potential, the corresponding input diode will conduct and the low voltage across the level shifting diodes will be insufficient to allow current to flow into the base of the transistor. Hence the transistor will be held in the cut-off state and the output will be high. With all three inputs in the high state, the level-shifting diodes will conduct and the resulting base current will bottom the transistor giving a low level output.

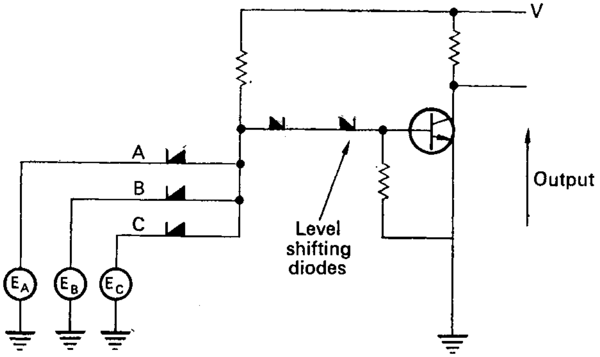


FIG. 11.12. Electronic form of NAND circuit.

11.1.5. The NOR *circuit*

The NOR circuit is one of the most widely used in logic circuits and can form the basis of a complete computer. It may be visualized by the circuit shown in Fig. 11.13.

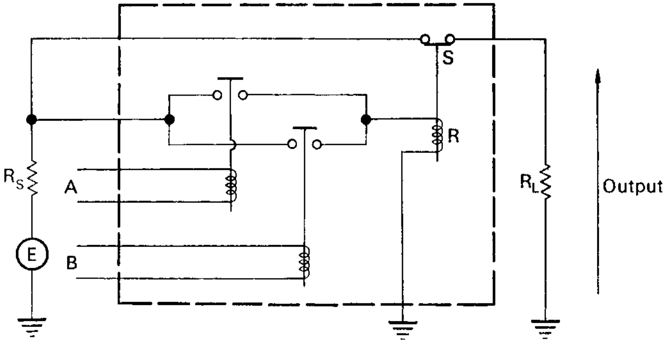


FIG. 11.13. Basic NOR circuit.

With no input on A and B the contacts S are closed and the output is in the high state of $ER_L/(R_S+R_L)$. If, however, either A or B are energized, the relay R will operate opening the contacts S and reducing the output to zero.

A	B	Output
0	0	1
0	1	0
1	0	0
1	1	0

FIG. 11.14. Truth table for NOR circuit.

The truth table for a two-input NOR circuit is shown in Fig. 11.14, and corresponds to an inversion of the output of an OR gate. In terms of Boolean algebra, the operation of the circuit may be described by

$$\text{NOT } (A \text{ or } B) = \text{NOT } A \text{ and NOT } B,$$

i.e.
$$\overline{A+B} = \bar{A}\cdot\bar{B}.$$

An electronic form of NOR circuit is indicated by Fig. 11.15. With both inputs at zero potential, the transistor will be held off

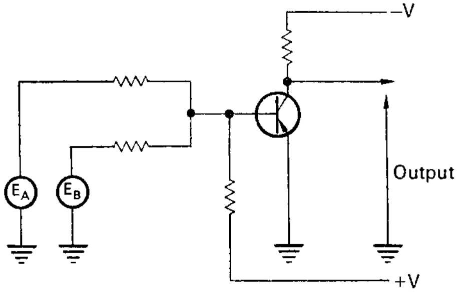


FIG. 11.15. Electronic form of NOR circuit.

by the positive bias on the base of the transistor and the output will be very close to $-V$. If a negative potential of sufficient magnitude is applied to either of the inputs the transistor will bottom and the output will approach a value of zero volts. This circuit is obviously an example of negative logic.

The NOR circuit (like the NAND circuit) can also be the basic unit of other circuits providing other logic function. If two NOR circuits are connected in series as shown in Fig. 11.16, the output of the first NOR gate will be $\overline{A+B}$ which, after inversion by the second NOR gate, gives $A+B$, i.e. an OR circuit.

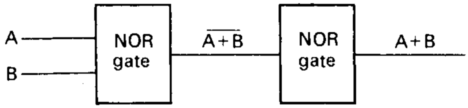


FIG. 11.16. OR circuit using NOR gates.

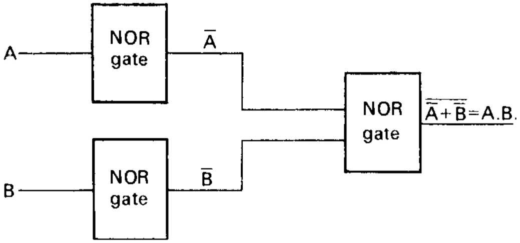


FIG. 11.17. AND circuit using NOR gate.

Similarly, with the combination shown in Fig. 11.17, an AND gate may be realized. A NOR gate in series with an input provides inversion, and the third NOR gate gives an output NOT (NOT A or NOT B), i.e. A and B ($= A{\cdot}B$).

11.1.6. *The exclusive* OR *circuit*

This circuit is only valid for two inputs and gives an output only when the inputs are different (i.e. $A = 0$ $B = 1$ or $A = 1$ $B = 0$). This means that the truth table will be as shown in Fig. 11.18, and a relay version of the exclusive OR circuit is as shown in Fig. 11.19. The inputs A and B are in opposition, and if they are of the same sign both pairs of relay contacts will remain open and the output will be at earth potential. If either one of the inputs is energized, one of the pairs of contacts will be closed and the output will rise to $ER_L/(R_S + R_L)$.

A	B	Output
0	0	0
0	1	1
1	0	1
1	1	0

FIG. 11.18. Truth table for exclusive OR circuit.

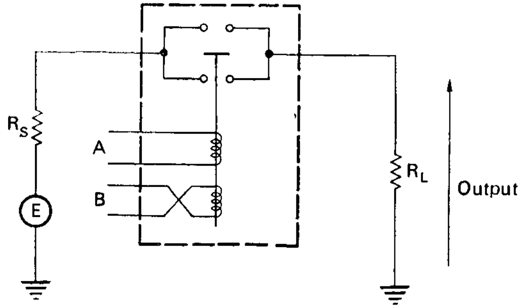


FIG. 11.19. Basic exclusive OR circuit.

In terms of Boolean algebra the exclusive OR circuit is equivalent to an output that is (A or B) and NOT (A and B),

i.e. $$\text{output} = (A+B)\cdot\overline{A\cdot B}.$$

A practical electronic circuit to realize the exclusive OR function would then involve several gates, and to avoid apparently complex circuitry this (and following circuits) will be shown in the form of a block diagram. From the expression given, an exclusive OR gate could be as illustrated by Fig. 11.20. The word "could" is used here since, as will be seen in the next section, alternative circuits may be used.

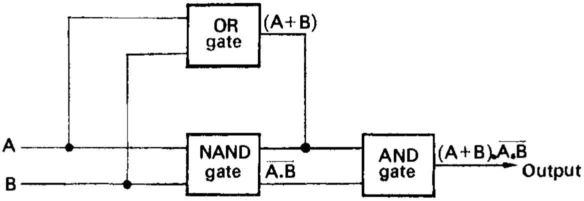


FIG. 11.20. Electronic form of exclusive OR circuit.

11.1.7. The half adder

Previous circuits described have in common the fact that only one output is required. The half adder has two outputs the need for which becomes apparent when the function of the circuit is considered in relation to binary addition.

Using binary arithmetic,

$$
\begin{aligned}
0+0 &= 0,\\
1+0 &= 1,\\
0+1 &= 1,\\
1+1 &= 0 \text{ and } 1 \text{ to carry.}
\end{aligned}
$$

These equations must not be confused with Boolean algebra or logic levels.

Reference to the truth table shown in Fig. 11.18 shows that the exclusive OR circuit will provide these results with the exception that the 1 to carry is omitted. The half adder provides both sum and carry outputs and may be built up by connecting an AND gate in parallel with an exclusive OR gate as shown in Fig. 11.21.

The truth table for the half adder is then as shown in Fig. 11.22.

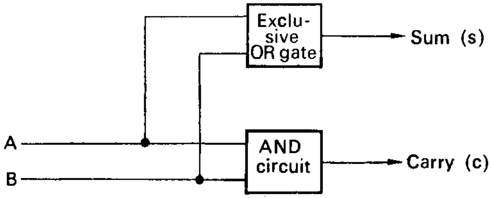


FIG. 11.21. The half adder.

A	B	S	C
0	0	0	0
0	1	1	0
1	0	1	0
1	1	0	1

FIG. 11.22. Truth table for half adder.

11.1.8. The full adder

This circuit is an extension of the half adder in as much as it can cope with three inputs and give two outputs representing the sum and carry terms in binary addition. As has been shown, a half adder provides outputs giving the sum and carry terms for two binary digits or bits. If the sum is passed to a second half adder it may be added to a third bit (which may be the carry input from a previous stage of addition), and the overall sum obtained. Both the carry terms may then be connected to the inputs of an OR gate to give the overall carry term.

With three inputs it is impossible to have high-level signals on both the carry outputs of the half adders, and the condition for both inputs to the OR gate being energized is not realized. The OR gate simply shows if one of its two inputs is energized. A full adder and the corresponding truth table are shown in Figs. 11.23 and 11.24.

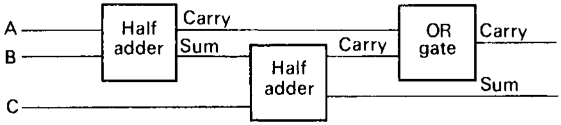


FIG. 11.23. The full adder.

A	B	C	Carry A+B+C	Sum A+B+C
0	0	0	0	0
0	0	1	0	1
0	1	0	0	1
0	1	1	1	0
1	0	0	0	1
1	0	1	1	0
1	1	0	1	0
1	1	1	1	1

FIG. 11.24. Truth table for full adder.

11.2. Boolean Algebra and Logic Design

In the middle of the nineteenth century, George Boole developed an algebra which is now an important tool in the analysis and design of digital switching systems. Where possible this has already been linked with the operation of circuits, and in this section the reduction of logic equations will be illustrated.

11.2.1. Rules of Boolean algebra

From the previous section the following rules have been discussed.

For the AND gate:

$$A \cdot A = A, \tag{11.2.1}$$

$$A \cdot 1 = A, \tag{11.2.2}$$

$$A \cdot 0 = 0, \tag{11.2.3}$$

$$A \cdot B = B \cdot A. \tag{11.2.4}$$

Also it does not matter in what order three inputs to an AND gate are combined,

i.e. $$A \cdot (B \cdot C) = (A \cdot B) \cdot C. \tag{11.2.5}$$

For the OR gate:

$$A + A = A, \tag{11.2.6}$$

$$A + 1 = 1, \tag{11.2.7}$$

$$A + 0 = A, \tag{11.2.8}$$

$$A + B = B + A. \tag{11.2.9}$$

The order in which the inputs to an OR gate are combined is irrelevant,

i.e. $$A + (B + C) = (A + B) + C. \tag{11.2.10}$$

For the NOT gate:

$$A \cdot \bar{A} = 0, \tag{11.2.11}$$

$$A + \bar{A} = 1. \tag{11.2.12}$$

For the NAND gate: $$\overline{A \cdot B} = \bar{A} + \bar{B}. \tag{11.2.13}$$

For the NOR gate: $$\overline{A + B} = \bar{A} \cdot \bar{B}. \tag{11.2.14}$$

A further relation involves two OR gates feeding an AND gate as shown in Fig. 11.25. A is common to both OR gates, and if it corresponds to 1 there will be an output 1. Alternatively, if A is 0 both B and C must be 1 for an output 1.

Therefore

$$(A + B) \cdot (A + C) = A + B \cdot C. \tag{11.2.15}$$

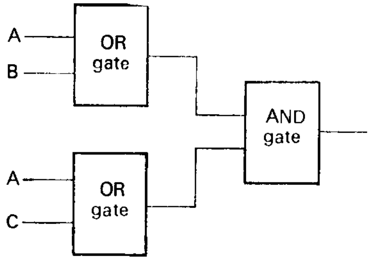


FIG. 11.25. Derivation of $(A+B)\cdot(B+C) = A+B\cdot C$.

If C is put equal to 0 in eqn. (11.2.15)

$$A\cdot(A+B) = A. \tag{11.2.16}$$

If C is put equal to A in eqn. (11.2.15),

$$(A+B)\cdot(A+A) = A+A\cdot B,$$

i.e. $$(A+B)\cdot A = A+A\cdot B$$

or $$A = A+A\cdot B. \tag{11.2.17}$$

In a similar manner,

$$(A\cdot B)+(A\cdot C) = A\cdot(B+C). \tag{11.2.18}$$

Although eighteen rules are postulated, and more may be derived, their use in logic design is fairly straightforward as the following examples show. It is useful to remember that there are equivalent rules for OR and AND, and in the absence of other information the AND function takes precedence over the OR function.

11.2.2. Examples on logic design

In section 11.1.6 the exclusive OR circuit was shown as a combination of an OR gate, a NAND gate, and an AND gate. The output was written in the form $(A+B)\cdot\overline{A\cdot B}$.

Hence, from eqn. (11.2.13),

$$\begin{aligned}\text{output} &= (A+B)\cdot(\bar{A}+\bar{B})\\ &= A\cdot\bar{A}+A\cdot\bar{B}+B\cdot\bar{A}+B\cdot\bar{B}.\end{aligned}$$

But from eqn. (11.2.11) $A\cdot\bar{A} = 0$ and $B\cdot\bar{B} = 0$.

Therefore $\qquad \text{output} = A\cdot\bar{B}+B\cdot\bar{A}.$

Hence an alternative means of deriving the exclusive OR function is as shown in Fig. 11.26.

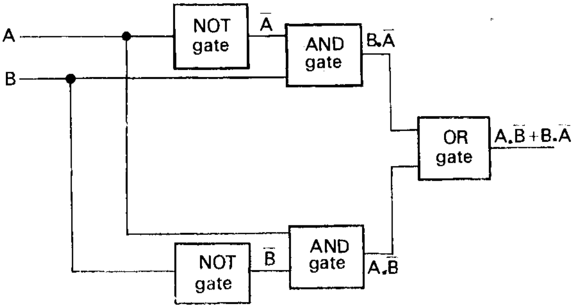


FIG. 11.26. Alternative exclusive OR circuit.

Alternatively, let $Z = A\cdot\bar{B}+B\cdot\bar{A}.$

Then $\qquad \bar{Z} = \overline{A\cdot\bar{B}+B\cdot\bar{A}}.$

But from eqn. (11.2.14),

$$\begin{aligned}\bar{Z} &= (\overline{A\cdot\bar{B}})\cdot(\overline{B\cdot\bar{A}})\\ &= (\bar{A}+B)\cdot(\bar{B}+A).\end{aligned}$$

Taking the inverse,

$$\begin{aligned}Z &= \overline{(\bar{A}+B)\cdot(\bar{B}+A)}\\ &= (\overline{\bar{A}+B})+(\overline{\bar{B}+A}).\end{aligned}$$

The exclusive OR circuit may then be drawn in terms of NOR gates as shown in Fig. 11.27.

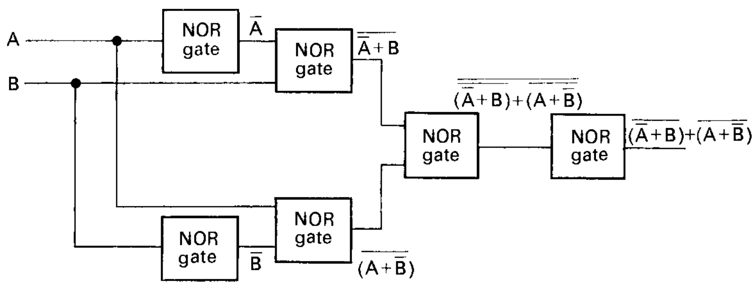


FIG. 11.27. Exclusive OR circuit using NOR gate.

As a further example, consider the full adder, the truth table for which is given in Fig. 11.24. From this table, by noting where the outputs are 1, the sum Z_1 may be written as Z_1 = (NOT A and NOT B and C) or (NOT A and B and NOT C) or (A and NOT B and NOT C) or (A and B and C),

i.e.
$$\begin{aligned} Z_1 &= \bar{A}\cdot\bar{B}\cdot C+\bar{A}\cdot B\cdot\bar{C}+A\cdot\bar{B}\cdot\bar{C}+A\cdot B\cdot C \\ &= C\cdot(\bar{A}\cdot\bar{B}+A\cdot B)+\bar{C}\cdot(\bar{A}\cdot B+A\cdot\bar{B}). \end{aligned}$$

But $C\cdot\bar{C}=0$, and it may easily be shown that $(\bar{A}\cdot\bar{B}+A\cdot B)\times(\bar{A}\cdot B+A\cdot\bar{B})=0$.

$$\begin{aligned} Z_1 &= (\bar{A}\cdot B+A\cdot\bar{B}+C)\cdot(\bar{A}\cdot\bar{B}+A\cdot B)+\bar{C}\cdot(\bar{A}\cdot B+A\cdot\bar{B}+C) \\ &= (\bar{A}\cdot B+A\cdot\bar{B}+C)(\bar{A}\cdot\bar{B}+A\cdot B+\bar{C}). \end{aligned}$$

Similarly, the carry term Z_2 may be written as:

Z_2 = (NOT A and B and C) or (A and B and NOT C)

or (A and C and NOT B) or (A and B and C)

$$\begin{aligned} &= \bar{A}\cdot B\cdot C+A\cdot B\cdot\bar{C}+A\cdot C\cdot\bar{B}+A\cdot B\cdot C \\ &= C(\bar{A}\cdot B+A\cdot\bar{B})+A\cdot B(C+\bar{C}) \\ &= C\cdot(\bar{A}\cdot B+A\cdot\bar{B})+A\cdot B, \end{aligned}$$

since
$$C+\bar{C}=1.$$

The full adder may then be drawn as shown in Fig. 11.28.

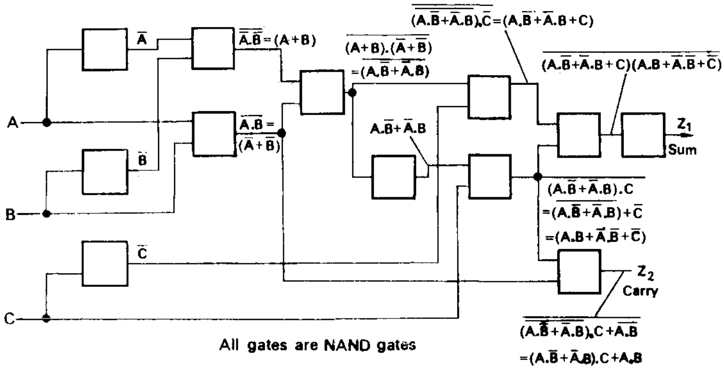

A
B
C
$\bar{A}$
$\bar{B}$
$\bar{C}$
$\overline{\bar{A}.\bar{B}}=(A+B)$
$\overline{A.B}=(\bar{A}+\bar{B})$
$\overline{\overline{(A.\bar{B}+\bar{A}.B)}.\bar{C}}=(A.\bar{B}+\bar{A}.B+C)$
$\overline{(A+B).(\bar{A}+\bar{B})}$
$=\overline{(A.\bar{B}+\bar{A}.B)}$
$\overline{(A.\bar{B}+\bar{A}.B+C)(A.B+\bar{A}.\bar{B}+\bar{C})}$
$A.\bar{B}+\bar{A}.B$
Z_1
Sum
$\overline{(A.\bar{B}+\bar{A}.B).C}$
$=\overline{(A.\bar{B}+\bar{A}.B)}+\bar{C}$
$=(A.B+\bar{A}.\bar{B}+\bar{C})$
Z_2
Carry
All gates are NAND gates
$\overline{\overline{(A.\bar{B}+\bar{A}.B).C}+\overline{A.B}}$
$=(A.\bar{B}+\bar{A}.B).C+A.B$

FIG. 11.28. Full adder using NAND gates.

11.3. Memory Circuits

Although the circuits previously discussed in this chapter enable logic functions to be obtained, none give a permanent indication of the result. A circuit that provides a memory is the bistable arrangement (see Volume 2, section 10.4), which in terms of logic circuits may be regarded as two NOR gates connected back to back as shown in Fig. 11.29.

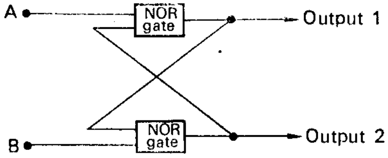


FIG. 11.29. Bistable circuit using NOR gate.

If pulses are applied to the inputs *A* and *B* in turn, the circuit will remember which was the last to be energized. Such a circuit is known as a RS (reset–set) bistable circuit, since one input (say *A*) receives the "set" pulses and *B* the "reset" pulses.

A variation, popular with integrated circuit manufacturers, is the J–K bistable circuit. In this arrangement the condition for the circuit to change its state is decided by the input conditions, but the actual change can only take place on receipt of a clock pulse applied to a third electrode.

11.4. Criteria of Logic Circuits

When comparing different types of logic circuit some criteria must be adopted. These are usually described as:

(1) fan-in;
(2) fan-out;
(3) noise margin;
(4) propagation delay.

11.4.1. Fan-in

This term describes the input circuit and is the number of inputs available. For the NAND circuit shown in Fig. 11.12 the fan-in is 3.

11.4.2. Fan-out

A logic circuit is rarely considered in isolation and is normally required to drive one or more other logic circuits. The fan-out is a measure of the number of circuits that may be successfully connected to the output.

11.4.3. Noise margin

The noise margin of a system is a measure of its immunity to noise. It is probably most easily visualized by reference to a simple inverting gate, a typical characteristic for which is shown by the full line in Fig. 11.30. The gate is assumed to remain in the high level V_1 for input voltages less than the switching voltage V_S and to switch to a low voltage V_2 for voltages greater than V_S. The dotted

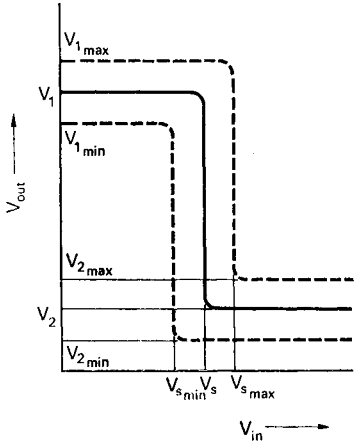
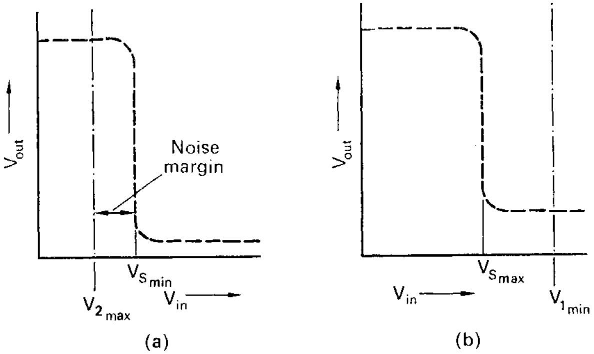


FIG. 11.30. Tolerances in switching characteristics.

lines indicate the possible variation of this characteristic due to manufacturing tolerances, differing temperatures, deviations of supply voltages, and varying load condition.

In normal operation, when one such gate drives a similar one, the worst possible conditions are illustrated by Fig. 11.31 (a) and (b). In Fig. 11.31 (a) the output of the first gate (i.e. the input to the second gate) is in the highest low voltage condition of $V_{2_{max}}$,



(a) (b)

FIG. 11.31. Noise margin.

and the switching voltage of the second gate has the minimum value of $V_{S_{\min}}$. The difference between $V_{S_{\min}}$ and $V_{2_{\max}}$ gives the margin between remaining in the high state and switching to the low state. Figure 11.31 (b) shows the output of the first gate to be in the lowest high voltage level (i.e. $V_{1_{\min}}$) and the second gate to have the highest switching voltage level of $V_{S_{\max}}$. The difference between $V_{1_{\min}}$ and $V_{S_{\max}}$ gives the safety margin between remaining in the low state and switching to the high state.

The values $(V_{S_{\min}}-V_{2_{\max}})$ and $(V_{1_{\min}}-V_{S_{\max}})$ are measures of the immunity of the gate from extraneous noise and the lower of these, normally $(V_{S_{\min}}-V_{2_{\max}})$ is known as the d.c. noise margin. Care must, however, be taken in using this value to compare systems since circuit impedances have a considerable effect, and the higher these are the greater is the vulnerability to interfering signals.

11.4.4. Propagation delay

Signals are delayed in passing through any circuit (in the simple relay circuits discussed this is the speed of operation of the relays) and the corresponding time, measured in microseconds or nano-

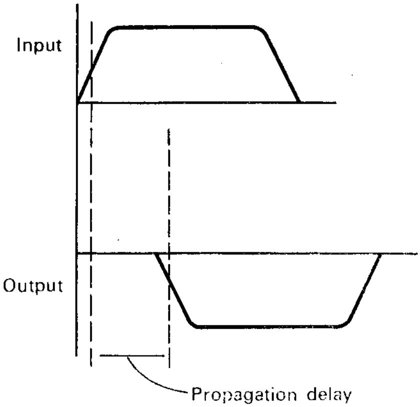


FIG. 11.32. Propagation delay.

seconds, is known as the propagation delay. This is measured between corresponding points on the input and output waveforms as indicated in Fig. 11.32.

11.5. Types of Logic Circuit

Over the years a variety of logic circuits have been developed and in this section some of the more common types will be discussed.

11.5.1. Direct-coupled transistor logic (DCTL)

This is one of the simplest types of logic circuit and, although now superseded, proved attractive to the early designers of integrated logic circuits. For a system based on positive logic, NOR and NAND circuits are as shown in Figs. 11.33 and 11.34. In practice the NOR circuit is more popular since the low level depends on the collector–emitter saturation voltage of these transistors in parallel, whereas in the NAND circuit it is the sum of three collector–emitter saturation voltages.

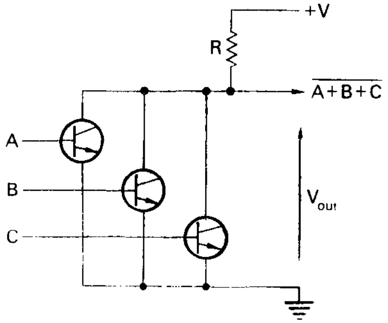


FIG. 11.33. DCTL NOR circuit.

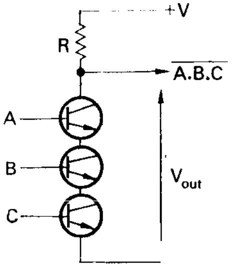


FIG. 11.34. DCTL NAND circuit.

Considering first the NOR circuit shown in Fig. 11.33; with all the inputs at zero the only currents flowing through the transistors will be their leakage currents, and the value of the output voltage will be decided by the load connected across the output terminals. This is the high state. When one of the inputs is a positive potential (i.e. in the high state) the corresponding transistor will be saturated and the additional current flowing through R will cause the output to drop to its low value. For the NAND circuit obviously all the inputs must be in the high state to give a low output.

The disadvantages of this type of logic circuit are the small difference between high- and low-level output voltages (0·8 and 0·2 V), a poor noise margin, and the effect known as "current hogging". When the gate drives similar circuits, the high-level voltage will be decided by the transistor (in the driven circuits) with the lowest base–emitter voltage. This transistor will tend to draw (hog) all the available current.

11.5.2. Resistor–transistor logic (RTL)

This type of logic circuit is a development of DCTL and incorporates resistors of about 500 Ω in series with the base connections. This has the effect of increasing the volt drop across the output terminals of a previous gate and increasing the high-level output. It also results in the output current being more equally shared between transistors.

As with all the circuits discussed, the propagation delay of RTL circuits is increased as the power dissipation is reduced. When the dissipation is about 12 mW the delay is about 12 ns but increases to 40 ns when the dissipation is reduced to 2 mW. Other typical characteristics of RTL circuits are high- and low-voltage levels of 2 and 1·5 V, a noise margin of 0·3 V, and a fan-out of 3 or 4.

11.5.3. Resistor–capacitor–transistor logic (RCTL)

This is a further development of DCTL with low-value capacitors (50 pF) connected across base resistors of about 20 kΩ. The collector load is also higher than the previous circuit (5 kΩ instead of

500 Ω), and this system is characterized by its higher impedances. Dissipations are low, about 2 mW per gate, and delay times are of the order of 150 ns.

11.5.4. Diode–transistor logic (DTL)

The basic circuit of this type of logic has already been described in section 11.1.4, and with positive logic this system is based on the NAND gate shown in Fig. 11.12. Alternatively, Fig. 11.12 may be used as a NOR gate with negative logic (i.e. the low level representing 1).

This logic system was one of the most widely used discrete forms in the early 1960's with a fan-out of 5 and typical logic levels of 4·5 and 0·3 V. The noise margin is in the range 0·4–0·9 V, and the propagation delay between 8 and 40 nS.

When an integrated circuit version of DTL was considered, difficulties arose due to the different requirements of the input and level-shifting diodes. These are ideally fabricated at the same time, but the former should be low storage devices (for a low propagation delay) while the latter should be high storage devices (for low turn-off times). This high turn-off time may be reduced by using an extra base power supply or reducing the base resistor, but the latter method leads to a reduction of base current and reduced fan-out capability.

11.5.5. Modified diode–transistor logic (MDTL)

This system arose out of the need to overcome the limited fan-out of DTL. The basic circuit is shown in Fig. 11.35, and the current gain of the first transistor Q_1 increases the base current of the second transistor Q_2, allowing a lower value of base resistor (5 kΩ instead of 20 kΩ). Typically, the fan-out of this circuit is 8.

The tapped resistor feeding the collector and base of Q_1 prevents

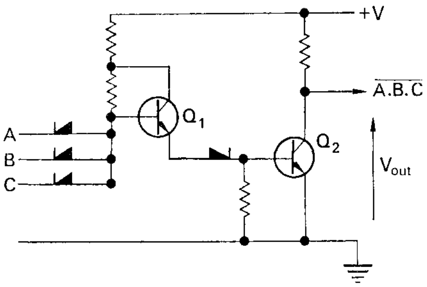


FIG. 11.35. Modified DTL circuit.

that transistor being saturated and leads to faster operation than DTL. The noise margin of MDTL is also slightly better than the basic DTL circuit.

11.5.6. Transistor–transistor logic (TTL or T^2L)

Although the circuit shown in Fig. 11.36 bears little resemblance to the DTL circuit shown in Fig. 11.12, the operation is very similar. The input diodes are replaced by junctions fabricated between the multi-emitters and base of transistor Q_1 and the level-shifting diodes by the base–collector junction of Q_1. The multi-emitter transistor itself is a circuit element peculiar to integrated circuits and is not available in discrete component form.

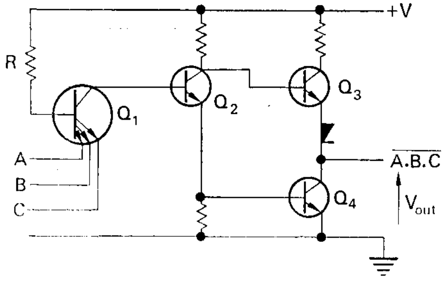


FIG. 11.36. TTL circuit.

As in the basic DTL circuit, Fig. 11.36 corresponds in positive logic to a NAND circuit. With all the inputs high, current flows into the base of Q_2, switching that transistor on. The fall in its collector voltage switches Q_3 off and the rise in its emitter voltage switches Q_4 on (the collector current of Q_4 being supplied by the following stage(s)). In this manner a low output is obtained. With one or more inputs in the low state, the current flowing through R is diverted through the base–emitter junction(s) and Q_2 is switched off, switching in turn Q_3 on and Q_4 off. The output is then in the high state. The diode ensures that when the "pull down" transistor Q_4 is on the "pull up", transistor Q_3 is off.

The main advantage of this type of circuit is the fast operation, since propagation delays are in the range 7–15 ns (for 25 and 10 mW dissipation respectively). Its other characteristics are about the same as those of conventional DTL circuits.

11.5.7. Complementary transistor logic (CTL)

Typical of this type of logic circuit is the AND gate shown in Fig. 11.37. With both the inputs in the high state, the transistors Q_1 and Q_2 are cut off, but Q_3 conducts to give a high output. With a low level input on either or both Q_1 and Q_2, the current flowing through the common collector resistor results in a the fall in voltage of the base of Q_3. This is sufficient to cut off the transistor.

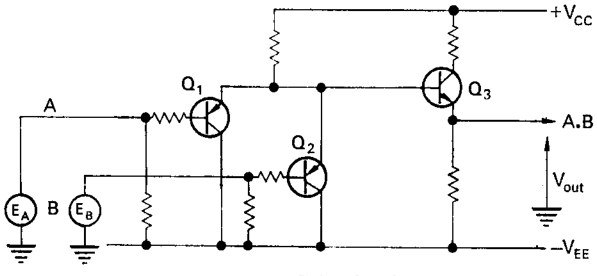


FIG. 11.37. CTL circuit.

Normally the dissipation of this circuit is relatively high (40 mW), but propagation delays of 4 ns may be obtained since the transistors are not saturated. Power supplies required are +4·5 V and −2 V, and the circuit gives logic levels of +2·5 V and −0·5 V.

11.5.8. Emitter-coupled or current mode logic (ECL)

As in the previous case, this type of logic circuit, shown in Fig. 11.38, uses transistors in the unsaturated condition. This enables low propagation delays to be obtained (6 ns) with a power dissipation of 35 mW.

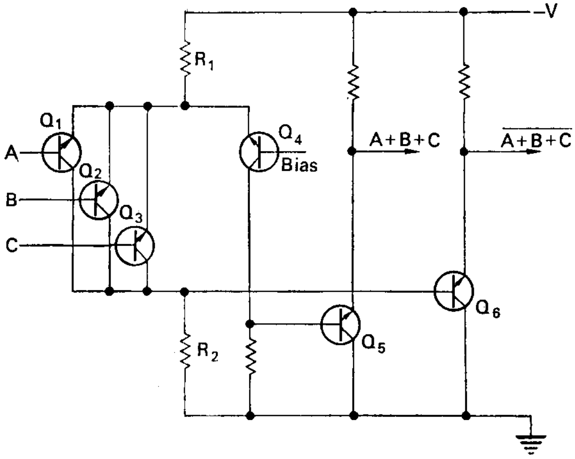


FIG. 11.38. ECL circuit.

When all the inputs are zero, Q_4 and Q_6 conduct and Q_5 is cut off, but if any input goes positive the voltage across R_1 will be sufficient to cut Q_4 off and give a high-level OR output. Also the voltage across R_2 will be sufficient to cut Q_6 off and give a low-level NOR output. The logic levels are about −0·75 V and −1·5 V with a noise margin of 0·3 V.

11.6. Choice of Logic Circuit

Where speed is the prime consideration, the non-saturation ECL type would be the obvious choice. Although the propagation delay of such a circuit is a few nanoseconds, it suffers from the disadvantages of relatively high power dissipation and low noise immunity.

Where speed is not essential, there is usually the choice between DTL, RTL, and TTL. Diode transistor logic is a useful general purpose logic circuit with a good noise margin. Resistor–transistor logic provides a good balance between propagation delay and power dissipation, but has a poor noise margin. Transistor–transistor logic is somewhat faster than DTL, but in other respects is about the same. The state at present suggests that TTL will be used for integrated circuit technology.

Questions on Chapter 11

Q.1. Sketch and explain the operation of the basic circuits used to perform the logic operations AND, OR, and NOT.

Q.2. Explain what is meant by NOR and NAND circuits and show how they may be used to perform the logic operation of AND, OR, and NOT.

Q.3. Explain with reference to the half adder how logic circuits can perform simple addition.

Q.4. Reduce to its simplest form the equation

$$(\bar{A}\cdot B+\bar{B}+B)\cdot(\bar{A}\cdot C\cdot C+\bar{B}).$$

Show how the result may be realized.

Q.5. Reduce to its simplest form the equation

$$K = A\cdot\bar{B}\cdot C+\bar{A}\cdot B\cdot C+\bar{A}\cdot B+\bar{B}\cdot C.$$

Show how the result may be realized.

Q.6. Discuss the criteria used to compare logic circuits.

Q.7. Both saturated and unsaturated transistors are used in logic circuits. Compare, with typical examples, the performances of such circuits.

CHAPTER 12

Electron Ballistics (VHF Valves)

THE subject of electron ballistics was introduced in Volume 1. The treatment is extended here with particular reference to VHF applications.

12.1. Motion in a Direct Electric Field

The simplest case of electron motion is when an electron is accelerated from rest under the influence of a voltage V. By the principle of conservation of energy, gain of kinetic energy = loss of potential energy,

i.e. $$\frac{mv^2}{2} = eV,$$

where m is the mass of an electron, v is the velocity attained, and e is the charge on an electron,

i.e. $$v = \sqrt{\left(\frac{2eV}{m}\right)}. \qquad (12.1.1)$$

This equation is valid for the majority of calculations, but at high velocities relativistic effects, i.e. the change in mass with velocity, cannot be ignored. This is based on Einstein's theory of relativity, the important relation for this analysis is that relating matter and energy,

i.e. $$W = mc^2, \qquad (12.1.2)$$

where W is the energy in joules, m is the mass in kilograms, and c is the velocity of electromagnetic waves in free space in metres per second.

Consider an electron being accelerated from rest through a voltage V,

$$\text{increase in energy} = \text{eV}.$$

The force increasing the velocity may be equated to the rate of change of momentum,

$$F = \frac{d}{dt}(mv).$$

If both m and v vary,

$$F = m\frac{dv}{dt}+v\frac{dm}{dt}. \tag{12.1.3}$$

But from eqn. (12.1.2), when this force acts through a distance dx,

$$F\,dx = dW = c^2\,dm,$$

i.e.

$$\begin{aligned} F &= c^2\frac{dm}{dx} \\ &= c^2\frac{dm}{dt}\frac{dt}{dx} \\ &= \frac{c^2}{v}\frac{dm}{dt}. \end{aligned} \tag{12.1.4}$$

From eqns. (12.1.3) and (12.1.4),

$$\frac{c^2}{v}\frac{dm}{dt} = m\frac{dv}{dt}+v\frac{dm}{dt},$$

i.e.

$$\left(\frac{c^2-v^2}{v}\right)\frac{dm}{dt} = m\frac{dv}{dt}$$

or

$$\frac{dm}{m} = \frac{v\,dv}{c^2-v^2},$$

integrating

$$\log_\varepsilon m = -\tfrac{1}{2}\log_\varepsilon(c^2-v^2)+K.$$

At $v = 0$, $m = m_0$ the rest mass.

Therefore $$K = \log_\varepsilon m_0 + \tfrac{1}{2}\log_\varepsilon c^2.$$

Therefore $$\log_\varepsilon m = -\tfrac{1}{2}\log_\varepsilon (c^2 - v^2) + \log_\varepsilon m_0 + \tfrac{1}{2}\log_\varepsilon c^2,$$

i.e. $$\log_\varepsilon \frac{m}{m_0} = \frac{1}{2}\log_\varepsilon \frac{c^2}{c^2 - v^2}$$

or $$\frac{m}{m_0} = \sqrt{\left[\frac{c^2}{(c^2 - v^2)}\right]},$$

i.e. $$m = \frac{m_0}{\sqrt{\left[\left(1 - \frac{v^2}{c^2}\right)\right]}}.$$

This expression shows that the effective mass of an electron increases with the velocity, theoretically approaching infinity at the speed of light.

But from eqn. (12.1.2)

$$mc^2 - m_0c^2 = eV.$$

Therefore $$\frac{m_0c^2}{\sqrt{\left(1 - \frac{v^2}{c^2}\right)}} - m_0c^2 = eV,$$

$$m_0c^2 = (m_0c^2 + eV)\sqrt{\left(1 - \frac{v^2}{c^2}\right)},$$

i.e. $$(m_0c^2)^2 = (m_0c^2 + eV)^2\left(1 - \frac{v^2}{c^2}\right)$$

or $$\frac{v^2}{c^2} = 1 - \frac{(m_0c^2)^2}{(m_0c^2 + eV)^2},$$

i.e. $$v = c\sqrt{\left(1 - \frac{1}{\left(1 + \frac{eV}{m_0c^2}\right)^2}\right)}.$$

This expression gives the true velocity of an electron when accelerated through a voltage V allowing for the change of mass. Substituting values and comparing this result with eqn. (12.1.1) shows that for a voltage less than about 5 kV, the error is negligible.

12.2. Motion in a Direct Electric Field Allowing for Space Charge

When the movement of a number of electrons is considered, the effect of the charge on the electrons themselves must be considered. The effect may be allowed for by means of Poisson's equation, which is derived from Gauss's theorem.

The latter states: *The surface integral of the normal component of the electric flux density over a closed surface (i.e. the total normal flux) is equal to the charge enclosed,*

i.e.
$$\oint_s \bar{D}\, ds = Q,$$

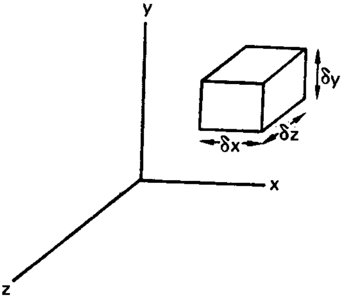


FIG. 12.1. Poisson's equation (rectangular coordinates).

i.e. for a small element δx, δy, δz shown in Fig. 12.1,

$$\frac{\partial D}{\partial x}\,\delta x\,\delta y\,\delta z+\frac{\partial D}{\partial y}\,\delta y\,\delta x\,\delta z+\frac{\partial D}{\partial z}\,\delta z\,\delta x\,\delta y = \varrho\,\delta x\,\delta y\,\delta z,$$

where ϱ is the charge density.

But $D = \varepsilon_0 E = -\varepsilon_0 \dfrac{\partial V}{\partial x} \left(\text{or } -\varepsilon_0 \dfrac{\partial V}{\partial y} \text{ or } -\varepsilon_0 \dfrac{\partial V}{\partial z} \right)$,

where ε_0 is the permittivity of free space.

Therefore
$$\varepsilon_0 \frac{\partial^2 V}{\partial x^2} + \varepsilon_0 \frac{\partial^2 V}{\partial y^2} + \varepsilon_0 \frac{\partial^2 V}{\partial z^2} = -\varrho.$$

When the only variation in V is along one axis (e.g. the y axis),
$$\frac{d^2 V}{dy^2} = -\frac{\varrho}{\varepsilon_0}.$$

This is Poisson's equation in one dimension.

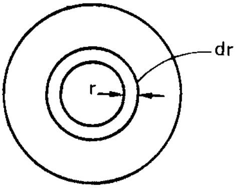


FIG. 12.2. Poisson's equation (cylindrical coordinates).

If cylindrical coordinates are considered as shown in Fig. 12.2, for an annular element of unit length:
$$d(2\pi r D) = \varrho 2\pi r\, dr,$$
$$\frac{1}{r}\frac{d}{dr}(rD) = \varrho,$$
or
$$\frac{D}{r} + \frac{dD}{dr} = \varrho.$$
But since
$$D = \varepsilon_0 E = -\varepsilon_0 \frac{dV}{dr},$$
$$\frac{1}{r}\frac{dV}{dr} + \frac{d^2 V}{dr^2} = -\frac{\varrho}{\varepsilon_0}.$$

This is Poisson's equation in cylindrical coordinates.

12.2.1. The planar diode

The simplest example of electron motion, allowing for space charge effects, is the planar vacuum diode shown in Fig. 12.3. In this analysis, relativistic effects are ignored, zero emission velocity assumed, and space charge conditions considered.

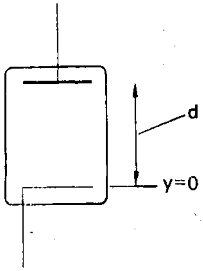


FIG. 12.3. The planar diode.

The current density J in a diode is equal to the charge flow per second per unit area,

i.e. $$J = -\varrho v,$$

where ϱ is the space charge density and v is the velocity.

But assuming zero emission velocity, the velocity of an electron that has moved through a voltage V is given by

$$eV = \frac{mv^2}{2},$$

i.e. $$v = \sqrt{\frac{2eV}{m}}.$$

Hence $$J = -\varrho\sqrt{\frac{2eV}{m}} \quad \text{or} \quad \varrho = -J\sqrt{\frac{m}{2eV}}.$$

Also from Poisson's equation,

$$\begin{aligned}\frac{d^2V}{dy^2} &= -\frac{\varrho}{\varepsilon_0}\\ &= \frac{J}{\varepsilon_0}\sqrt{\frac{m}{2eV}},\end{aligned}$$

where y is measured from the cathode surface as shown in Fig. 12.3.

Multiplying by $2\dfrac{dV}{dy}$ and integrating,

$$\left(\frac{dV}{dy}\right)^2 = \frac{4J}{\varepsilon_0}\left(\sqrt{\frac{m}{2e}}\right)V^{1/2}+K.$$

To determine the constant K, the conditions at the cathode surface are considered. At this point the positive field due to the anode voltage is balanced by the negative field due to the space charge itself,

i.e. at $\quad y = 0, \quad V = 0, \quad$ and $\quad \dfrac{dV}{dy} = 0.$

Hence $\quad K = 0.$

Therefore $$\frac{dV}{dy} = \sqrt{\left[\frac{4J}{\varepsilon_0}\left(\sqrt{\frac{m}{2e}}\right)V^{1/2}\right]}.$$

Rearranging and integrating,

$$\frac{4V^{3/4}}{3} = \sqrt{\left[\frac{4J}{\varepsilon_0}\left(\sqrt{\frac{m}{2e}}\right)\right]}\,y+K_1.$$

Using the same initial condition, i.e. at $y = 0$, $V = 0$, then $K_1 = 0$

Therefore $$\frac{4}{3}V^{3/4} = \sqrt{\left[\frac{4J}{\varepsilon_0}\left(\sqrt{\frac{m}{2e}}\right)\right]}\,y$$

or $$J = \frac{4\varepsilon_0}{9}\left(\sqrt{\frac{2e}{m}}\right)\frac{V^{3/2}}{y^2}.$$

The anode current density may be derived by putting y equal to the anode–cathode spacing d and V equal to V_A the anode voltage,

i.e.
$$J = \frac{4\varepsilon_0}{9}\left(\sqrt{\frac{2e}{m}}\right)\frac{V_A^{3/2}}{d^2}\,.$$

Substituting values for ε_0, e and m gives

$$J = 2{\cdot}33\times 10^{-6}\,\frac{V_A^{3/2}}{d^2}\quad \text{A/m}^2.$$

12.2.2. The coaxial diode

A more practical case than the planar diode is the coaxial diode shown in Fig. 12.4. The element dr at a radius r is considered and a line source is assumed.

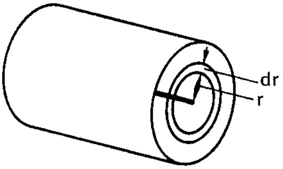


FIG. 12.4. The coaxial diode.

Current per unit length of emitter, $I = -2\pi r\varrho v$ *A/m*, where ϱ is the space charge density and v is the velocity.

Also, as in the previous case,

$$v = \sqrt{\frac{2eV}{m}}\,.$$

Therefore
$$I = -2\pi r\varrho\sqrt{\left(\frac{2eV}{m}\right)}$$

or
$$\varrho = -\frac{I}{2\pi r}\sqrt{\left(\frac{m}{2eV}\right)}.$$

The cylindrical form of Poisson's equation is

$$\frac{1}{r}\frac{dV}{dr}+\frac{d^2V}{dr^2}=\frac{I}{2\pi r\varepsilon_0}\sqrt{\frac{m}{2eV}}\,.$$

The solution of this equation could be of the same form as that for a planar diode. Hence assuming a solution $V = Kr^n$, where K is a constant,

$$\frac{dV}{dr}=nKr^{n-1},$$

$$\frac{d^2V}{dr^2}=n(n-1)Kr^{n-2}.$$

Therefore $$\frac{1}{r}nKr^{n-1}+n(n-1)Kr^{n-2}=\frac{I}{2\pi r\varepsilon_0}\sqrt{\frac{m}{2eKr^n}}\,,$$

i.e. $$n^2Kr^{n-2}=\frac{I}{2\pi r\varepsilon_0}\sqrt{\frac{m}{2eKr^n}}$$

or $$K^{3/2}=\frac{I}{2\pi r^{[(3n/2)-1]}\varepsilon_0}\sqrt{\frac{m}{2e}}\,.$$

r may be eliminated from this expression by assuming $n = \frac{2}{3}$.

Then $$K^{3/2}=\frac{I}{2\pi\varepsilon_0}\sqrt{\frac{m}{2e}}$$

or $$K^3=\frac{I^2m}{8\pi^2\varepsilon_0^2e}\,.$$

Hence $$V=\left(\frac{I^2m}{8\pi^2\varepsilon_0^2e}\right)^{1/3}r^{2/3}$$

is a solution

or $$J=\left[2(\sqrt{2})\pi\varepsilon_0\sqrt{\frac{e}{m}}\right]\frac{V^{3/2}}{r}$$

$$=14{\cdot}7\times10^{-4}\frac{V^{3/2}}{r}\quad\text{A/m.}$$

The anode current density is obtained by putting $r = r_A$ the anode radius,

i.e. $$J = 14{\cdot}7 \times 10^{-4} \frac{V^{3/2}}{r_A} \quad \text{A/m.}$$

This expression is in the form of a "three-halves power law".

12.2.3. Other vacuum valves

Two particular devices have been analysed and, in general, all vacuum diodes operated under space charge conditions will obey a law $I = kV^{3/2}$. When emission velocities are allowed for, evaluation of the equations are made more difficult by a finite value of the potential gradient at the cathode and a constant appearing in the expression for the velocity $\left(\text{i.e. } v = \sqrt{\frac{2e}{m}(V+K)}\right)$. This leads to the idea of a virtual cathode just above the emitting surface and a reduction in the effective cathode–anode distance.

The introduction of a grid (i.e. a triode) leads to the effective accelerating voltage acting on the electrons being due to a combination of grid voltage V_G and anode voltage V_A. Since the grid is that much closer, its effect will be greater, and to a close approximation the equation may be written

$$I_A = K\left(V_G + \frac{V_A}{\mu}\right)^{3/2}$$

where μ is a constant for the triode, actually the amplification factor.

12.3. Deflecting Effect of an Electric Field

The best illustration of the deflecting effect of an electric field is the deflecting system of an electrostatic cathode-ray tube. The first part of this was covered in detail in Volume 1, but it will be repeated briefly here.

If an electron accelerated through a potential V_A enters the electric field normally with a velocity v_x as shown in Fig. 12.5,

$$v_x = \sqrt{\frac{2eV_A}{m}},$$

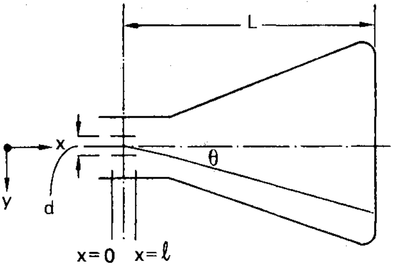


FIG. 12.5. Electrostatic deflexion.

and within the field, i.e. $0 < x < l$,

$$x = \left(\sqrt{\frac{2eV_A}{m}}\right)t. \tag{12.3.1}$$

The deflecting force eV_D/d gives rise to an acceleration in the y direction of eV_D/md,

i.e.
$$y = \frac{1}{2}\left(\frac{eV_D}{md}\right)t^2. \tag{12.3.2}$$

Eliminating t between eqns. (12.3.1) and (12.3.2) gives

$$\begin{aligned} y &= \frac{1}{2}\left(\frac{eV_D}{md}\right)\frac{x^2}{(2eV_A/m)} \\ &= \frac{1}{4}\frac{V_D}{V_A}\frac{x^2}{d}, \end{aligned}$$

i.e. electrons follow a parabolic path between the plates. Also the angle of exit is given by

$$\left(\frac{dy}{dx}\right)_{x=l} = \frac{1}{2}\frac{V_D}{V_A}\frac{l}{d}$$
$$= \tan\theta.$$

Hence the vertical displacement when the electrons leave the deflecting plates is

$$y = \frac{1}{4}\frac{V_D}{V_A}\frac{l^2}{d}$$
$$= \frac{l}{2}\tan\theta,$$

i.e. electrons may be regarded as being emitted in a straight line from a point at the centre of the deflecting system. Hence if the distance between the centre of the plates and the screen is L, the sensitivity, defined as the deflexion per unit deflecting voltage, is given by

$$\text{sensitivity} = \frac{L\tan\theta}{V_D}$$
$$= \frac{Ll}{2V_A d}.$$

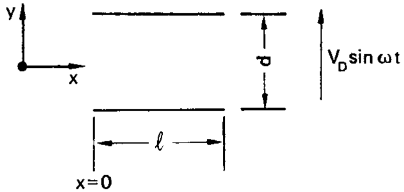


FIG. 12.6. Dynamic deflexional sensitivity.

When a very high frequency, deflecting signal is considered, the transit time of electrons must also be allowed for. Assume the electron enters the field $\frac{V_D}{d}\sin\omega t$ as shown in Fig. 12.6, where V_D

is the peak value of the deflecting voltage and d the distance between the plates, with a velocity $v_x = \sqrt{\dfrac{2eV_A}{m}}$.

Since the initial velocity in the vertical direction is zero, after a time t_1 the velocity is $v_y = \int_t^{t+t_1} f\,dt$, where $f = \dfrac{eV_D \sin \omega t}{md}$ and t_1 is the time taken to move a distance x $\left(\text{i.e. } t_1 = \dfrac{x}{v_x} = x\sqrt{\dfrac{m}{2eV_A}}\right)$.

Therefore
$$\begin{aligned} v &= \int_t^{t+t_1} \frac{eV_D \sin \omega t}{md}\,dt \\ &= \frac{eV_D}{md}\left[-\frac{\cos \omega t}{\omega}\right]_t^{t+t_1} \\ &= \frac{eV_D}{\omega md}[\cos \omega t - \cos \omega(t+t_1)]. \end{aligned}$$

When $x = l$, the length of deflecting plates

$$t_1 = l\sqrt{\frac{m}{2eV_A}}.$$

Hence an electron leaves the field at an angle θ, where

$$\begin{aligned} \tan\theta &= \frac{v_y}{v_x} \\ &= \frac{V_D}{\omega d}\left(\sqrt{\frac{e}{2mV_A}}\right)[\cos \omega t - \cos \omega(t+t_1)] \\ &= \frac{V_D}{\omega d}\left(\sqrt{\frac{e}{2mV_A}}\right) 2 \sin \omega\left(t+\frac{t_1}{2}\right) \sin \frac{\omega t_1}{2}. \end{aligned}$$

And assuming the electron may be regarded as being emitted from a point distance L from the screen.

Deflexion on screen $= L \tan \theta$

$$= L\frac{V_D}{\omega d}\left(\sqrt{\frac{e}{2mV_A}}\right) 2 \sin \omega\left(t+\frac{t_1}{2}\right) \sin \frac{\omega t_1}{2}.$$

This has a peak value of $\frac{LV_D}{\omega d}\left(\sqrt{\frac{e}{2mV_A}}\right)2 \sin \frac{\omega t_1}{2}$, giving a dynamic, deflexional sensitivity

$$= \frac{2L}{\omega d}\left(\sqrt{\frac{e}{2mV_A}}\right) \sin \frac{\omega t_1}{2}.$$

The dynamic, deflexional sensitivity falls to zero when

$$\frac{\omega t_1}{2} = \pi,$$

i.e.
$$f = \frac{1}{t_1} = \frac{1}{l}\sqrt{\left(\frac{2eV_A}{m}\right)}.$$

If $V_A = 1$ kV, $\frac{e}{m} = 1{\cdot}76\times 10^{11}$ C/kg, $l = 1$ cm, this corresponds to a frequency $\frac{1}{10^{-2}}\sqrt{(2\times 1{\cdot}76\times 10^{11}\times 10^3)}$, i.e. 1870 MHz.

12.4. Focusing Effect of an Electric Field

Previous sections have considered the accelerating or deflecting effect of an electric field. In a cathode ray tube it is necessary to accelerate electrons so that they strike the screen with sufficient energy to illuminate the screen and also focus them to produce a spot, say 1/50 in. in diameter. The latter may be accomplished either by electrostatic or electromagnetic means, and in this section the operation of an "electrostatic lens" will be discussed.

12.4.1. Refraction by an electric field

Consider two parallel equipotential surfaces at potentials V_1 and V_2 as shown in Fig. 12.7.

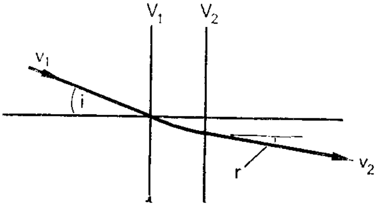


FIG. 12.7. Refraction by an electric field.

Assume an electron has initially started from rest. When it reaches the equipotential surface V_1 its velocity (v_1) is given by $v_1 = \sqrt{\frac{2eV_1}{m}}$ and when it reaches V_2 its velocity will be $v_2 = \sqrt{\frac{2eV_2}{m}}$.

When the electron travels in the direction of the field it will simply be accelerated. If its path makes an angle of incidence i with the electric field, the electron will be accelerated in the direction of the field, but the component of velocity at right angles to the field will be unaltered.

Hence the electron will leave at an angle r where

$$v_1 \sin i = v_2 \sin r$$

or

$$\left(\sqrt{\frac{2eV_1}{m}}\right) \sin i = \left(\sqrt{\frac{2eV_2}{m}}\right) \sin r$$

or

$$\frac{\sin i}{\sin r} = \sqrt{\frac{V_2}{V_1}}.$$

The ratio $\sqrt{(V_2/V_1)}$ corresponds to the refractive index of the electric field. In fact, by an argument somewhat similar to that used in the previous section, the path of the electron between the equipotential surfaces may be shown to be parabolic. If V_2 is greater than V_1, the divergence of the electron will be reduced.

12.4.2. *Electron lenses*

The previous subsection has indicated the possibility of refraction by an electric field. Consider first the elemental lens shown in Fig. 12.8. If D_z and $\left(D_z+\frac{\partial D_z}{\partial z}\,\delta z\right)$ is the electric flux density in an axial direction on each side of the element and D_r the radial flux

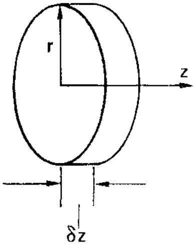


FIG. 12.8. Elemental electrostatic lens.

density at a radius r, by Gauss's theorem the total normal flux is equal to the charge enclosed,

i.e.
$$\pi r^2\left(D_z+\frac{\partial D_z}{\partial z}\,\delta z\right)+2\pi r\,\delta z D_r-\pi r^2 D_z = 0,$$

i.e.
$$r\frac{\partial D_z}{\partial z}+2D_r = 0$$

or
$$D_r = -\frac{r}{2}\,\frac{\partial D_z}{\partial z}.$$

But
$$D_r = \varepsilon_0 E_r \quad \text{and} \quad D_z = \varepsilon_0 E_z.$$

Therefore
$$E_r = -\frac{r}{2}\,\frac{\partial E_z}{\partial z}.$$

The force on diverging electrons is then

$$-eE_r\left(=\frac{re}{2}\frac{\partial E_z}{\partial z}\right)$$

and if $\partial E_z/\partial z$ is negative, this force, being proportional to the radius, tends to produce a point focus.

Therefore

$$m\frac{\partial^2 r}{\partial r^2}=\frac{re}{2}\frac{\partial E_z}{\partial z}$$

or
$$m\frac{\partial^2 r}{\partial r^2}=-\frac{re}{2}\frac{\partial^2 V_z}{\partial z} \qquad (12.4.1)$$

since
$$E_z=-\frac{\partial V_z}{\partial z^2}\,.$$

Consider now the three-electrode lens shown in Fig. 12.9. The axial movement is described by the equation

$$m\frac{\partial^2 z}{\partial r^2}=e\frac{\partial V}{\partial z}\,,$$

and from the voltage distribution curve it follows that an electron entering with a velocity v_0 is initially decelerated, the velocity falling to a minimum at the centre of the lens. It then accelerates until it leaves the lens with a velocity v_0, i.e. the same as the initial velocity.

The radial movement is described by eqn. (12.4.1),

i.e.
$$m\frac{\partial^2 r}{\partial r^2}=-\frac{re}{2}\frac{\partial^2 V}{\partial z^2}\,,$$

and when $\partial^2 V/\partial z^2$ is positive, diverging electrons will be accelerated towards the centre line of the lens. Since the electron spends more

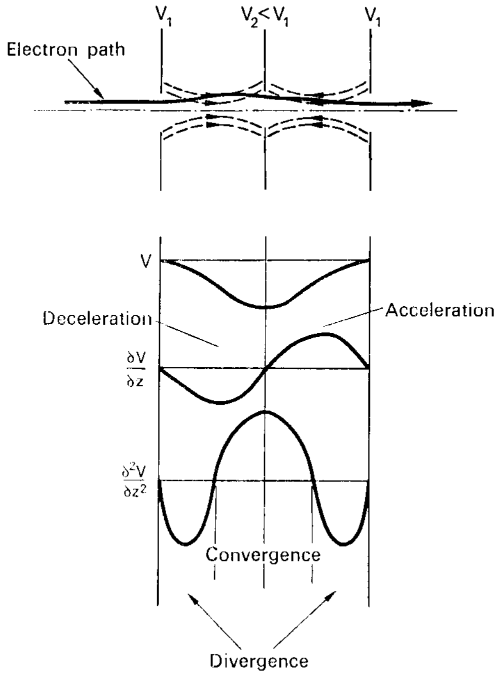


FIG. 12.9. Three-electrode lens.

time in the central (converging) region, of the three-electrode lens shown, than the outer (diverging) regions, the net effect will be to produce a point focus for a diverging electron beam.

12.5. Combined Effect of Parallel Direct and Alternating Electric Fields

In most of the previous sections the effects of direct fields have been considered. In general, when combined direct and alternating fields are considered, the transit time is of prime importance. For triodes one of the effects of transit time is the reduction of input impedance discussed in Chapter 9, while in klystrons the transit time is an inherent part of the operation, Consider, first, the structure shown in Fig. 12.10. The two grids G_1 and G_2 are maintained at a d.c. potential V_0 and have an alternating potential $\delta V_0 \sin \omega t$ between them.

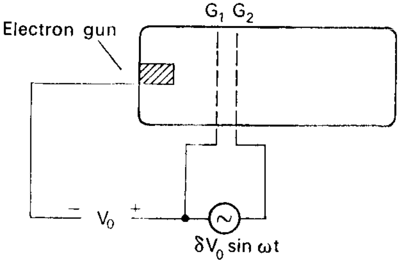


FIG. 12.10. Velocity modulation.

Ignoring emission velocities and the effects of space charge, the speed of electrons leaving G_2 is given by

$$\tfrac{1}{2}mv^2 = e(V_0 + \delta V_0 \sin \omega t)$$

or

$$v = \sqrt{\left[\frac{2eV_0}{m}\left(1 + \frac{\delta V_0}{V_0}\sin \omega t\right)\right]}$$

$$= v_0\left(1 + \frac{\delta V_0}{V_0}\sin \omega t\right)^{1/2},$$

where v_0 is the velocity of the electrons leaving when $\sin \omega t = 0$.

But if $\dfrac{\partial V_0}{V_0}\sin \omega t$ is less than about $\frac{1}{10}$, the binomial expansion may be used and high order terms omitted,

i.e.

$$v = v_0\left(1 + \frac{1}{2}\,\frac{\delta V_0}{V_0}\sin \omega t\right).$$

Hence the electrons leave with velocities ranging from $v_0 - \delta v_0$ to $v_0 + \delta v_0$, where $\delta v_0 = \dfrac{v_0}{2}\,\dfrac{\delta V_0}{V_0}$. If the space beyond G_2 has no electric field (i.e. a drift space), at some point the electrons will tend to bunch, the faster electrons catching up the slower ones. This is best illustrated by the Applegate diagram shown in Fig. 12.11.

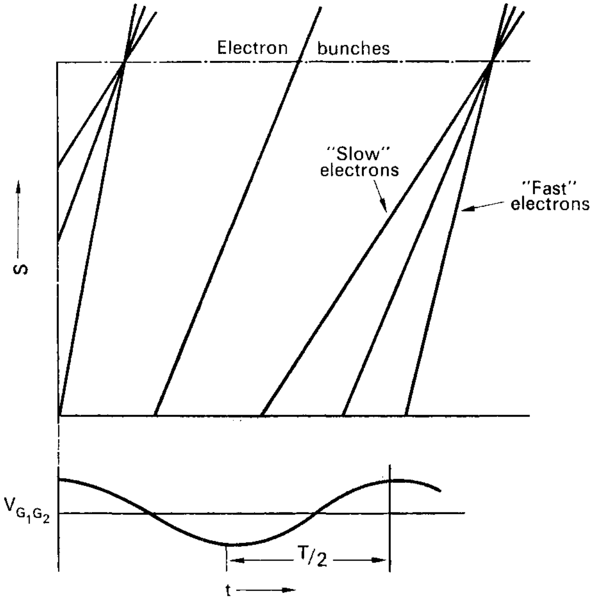


FIG. 12.11. Applegate diagram.

The distance to the first bunch s corresponds to the time taken for the fastest electrons to catch the slowest,

i.e. $$\frac{s}{v_0-\delta v_0}-\frac{s}{v_0+\delta v_0}=\frac{T}{2},$$

where T is the period,

i.e. $$\frac{2s\,\delta v_0}{v_0^2-(\delta v_0)^2}=\frac{T}{2}.$$

Since $(\delta v_0)^2 \ll v_0^2$,

$$s = \frac{Tv_0^2}{4\delta v_0}$$

$$= \frac{v_0^2}{4f\,\delta v_0}.$$

But $v_0 = \sqrt{\frac{2eV_0}{m}}$ and $\delta v_0 = \frac{v_0}{2}\,\frac{\delta V_0}{V_0}$,

i.e.

$$s = \frac{\sqrt{\left(\frac{2eV_0}{m}\right)}}{4f\frac{1}{2}\,\frac{\delta V_0}{V_0}}$$

$$= \frac{1}{(\sqrt{2})f}\left(\sqrt{\frac{e}{m}}\right)\frac{V_0^{3/2}}{\delta V_0}.$$

If a second pair of grids and a collecting electrode are now included, the electrode structure becomes as shown in Fig. 12.12.

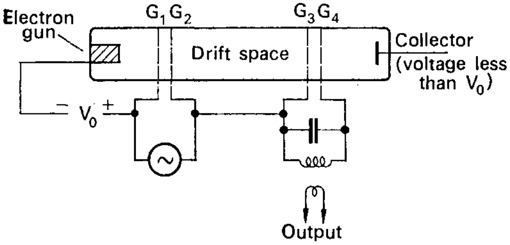


FIG. 12.12. The klystron.

The second pair of grids G_3 and G_4 are arranged such that the electrons have had time to bunch. These bunches pass through the grids G_3G_4 in the process inducing a voltage between them, and pulsing the tuned output circuit.

In a practical klystron, tuned circuits would be used for both the input and output circuits, these being in the form of cavity resonators as shown in Fig. 12.13. These cavities are known as the buncher and catcher respectively.

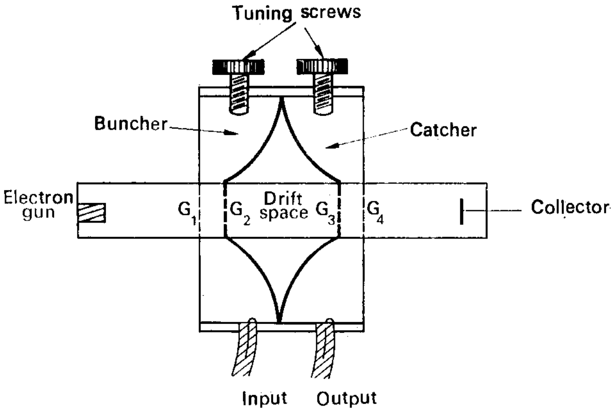


FIG. 12.13. Practical klystron.

Klystrons operate in the range 100–10,000 MHz, typically giving a power output of 10 W at 3000 MHz with a power gain of 20 dB. Multi-cavity klystrons with four cavities are also available, giving continuous wave powers of up to 500 kW.

A klystron may be used as an oscillator by coupling the output cavity to the input cavity by a line of such length as to provide positive feedback. This, in fact, corresponds to a phase change of $(3\pi)/2$ or $(7\pi)/2$, etc. It is more convenient, however, to use the same cavity for the buncher and catcher as shown in Fig. 12.14. This is known as a reflex klystron.

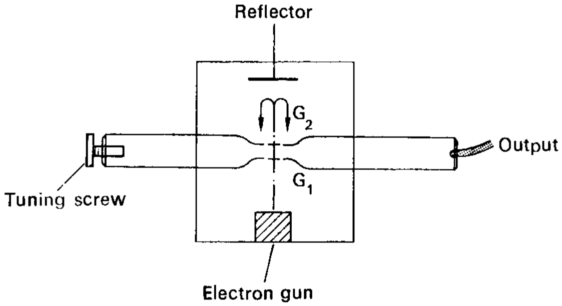


FIG. 12.14. The reflex klystron.

Assume the device is oscillating with an alternating voltage between G_1G_2. Electrons, emitted from the gun, pass through these grids and are velocity-modulated. They then form bunches in the space above G_2. After reflection, due to the negative potential on the reflector, the bunches will again pass through the grids and, under operating conditions, will sustain the oscillations. Typically, the cavity would be at earth potential, the cathode between 400 V and 1 kV negative with respect to the resonator, and the reflector between 100 and 300 V more negative than the cathode.

The frequency, to which the resonator must be tuned, may be estimated by knowing that the phase shift between leaving and re-entering G_2 must be $(3\pi)/2$, $(7\pi)/2$, etc. For the electron leaving at the instant when the alternating voltage is zero, the velocity on leaving G_2 is given by $v = \sqrt{\frac{2eV_1}{m}}$, where V_1 is the direct voltage between resonator and cathode. The retarding force in the drift space is eV_2/d, where V_2 is the direct voltage between cavity and reflector and d the distance between G_2 and reflector.

Hence the retardation is $(eV_2)/(md)$ and the time taken between leaving G_2 and coming to rest is given by

$$0 = \sqrt{\left(\frac{2eV_1}{m}\right)} - \frac{eV_2}{md}t,$$

i.e.

$$t = \frac{d}{V_2}\sqrt{\left(\frac{2mV_1}{e}\right)},$$

and the total time between leaving and re-entering G_2 is

$$2\frac{d}{V_2}\sqrt{\left(\frac{2mV_1}{e}\right)}.$$

But this corresponds to $\frac{3}{4f}$, $\frac{7}{4f}$, etc.,

i.e.
$$\frac{2d}{V_2}\sqrt{\frac{2mV_1}{e}} = \frac{3}{4f},\quad \frac{7}{4f},\ \text{etc.},$$

i.e.
$$f = \frac{3V_2}{8d}\sqrt{\left(\frac{e}{2mV_1}\right)},\quad \frac{7V_2}{8d}\sqrt{\left(\frac{e}{2mV_1}\right)},\quad \text{etc.}$$

Typical reflex klystrons, such as those used for microwave test benches, provide an output of a few watts with an efficiency of about 5%. Although several values of reflector voltage are possible for a given frequency, maximum power is normally attained with the reflector at its minimum operating voltage.

12.6. Deflecting Effect of a Magnetic Field

A magnetic field has no effect on an electron at rest and it can only be considered in relation to a moving electron. Even under these conditions, the force exerted is always at right angles to the direction of motion and no work is done, so that the magnitude of the velocity is unaltered. This means that a magnetic field can only deflect an electron, the deflecting force being given by

$$F = Bev,$$

where F is the force (newtons), B is the magnetic field (tesla), e is the charge on an electron (coulomb), and v is the velocity (m/s).

The direction of this force is at right angles to both the field and the velocity.

Two examples will be considered here, deflexion in a cathode-ray tube and a diode with a transverse field. The latter leads to the magnetron principle.

12.6.1. Deflexion in a cathode-ray tube

As stated, an electron entering a magnetic field moves with constant velocity and is subjected to a force at right angles to the direction of motion. The path of a particle under such conditions must be the arc of a circle as indicated in Fig. 12.15.

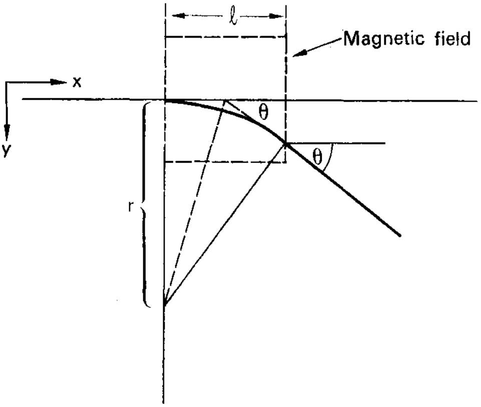


FIG. 12.15. Electromagnetic deflexion.

In a cathode-ray tube, electrons are initially accelerated by the final anode to a velocity $v_x = \sqrt{\frac{2eV_A}{m}}$. When a magnetic deflecting system is incorporated, they move in the arc of a circle the radius of which may be found by equating the force due to the magnetic field with that due to the centrifugal force,

i.e. $$Bev_x = \frac{mv_x^2}{r},$$

i.e. $$r = \frac{mv_x}{Be} = \frac{1}{B}\sqrt{\left(\frac{2mV_A}{e}\right)}.$$

Hence the equation of motion, taking the point of entry as the origin, is given by

$$x^2+\left[y-\frac{1}{B}\sqrt{\left(\frac{2mV_A}{e}\right)}\right]^2 = \frac{2mV_A}{B^2e}.$$

The angle at which the electron leaves the field is obtained by differentiating this expression and putting $x = l$,

i.e. $$2x+2\left[y-\frac{1}{B}\sqrt{\left(\frac{2mV_A}{e}\right)}\right]\frac{dy}{dx} = 0,$$

i.e. $$\frac{dy}{dx} = -x\bigg/\left[y-\frac{1}{B}\sqrt{\left(\frac{2mV_A}{e}\right)}\right].$$

But when $x = l$

$$\left[y-\frac{1}{B}\sqrt{\left(\frac{2mV_A}{e}\right)}\right]^2 = \frac{2mV_A}{B^2e}-l^2.$$

Taking the square root and substituting

$$\frac{dy}{dx} = \pm l\bigg/\left[\sqrt{\left(\frac{2mV_A}{B^2e}-l^2\right)}\right].$$

The positive value must be chosen giving

$$\frac{dy}{dx} = 1\bigg/\left[\sqrt{\left(\frac{2mV_A}{B^2el^2}-1\right)}\right].$$

Normally, $$\frac{2mV_A}{B^2el^2} \gg 1.$$

Therefore $$\frac{dy}{dx} = Bl\sqrt{\left(\frac{e}{2mV_A}\right)},$$

to a first approximation. The electron may be regarded as emanating from a point at the centre of the field, and if this is at a distance L from the screen the deflexion on the screen is given by

$$D = BlL\sqrt{\left(\frac{e}{2mV_A}\right)},$$

i.e. $$\text{sensitivity} = lL\sqrt{\left(\frac{e}{2mV_A}\right)} \quad \text{m/T}.$$

12.6.2. Planar diode with transverse magnetic field

In this case an electron, initially at rest, is accelerated by the electric field and at the same time deflected by the magnetic field.

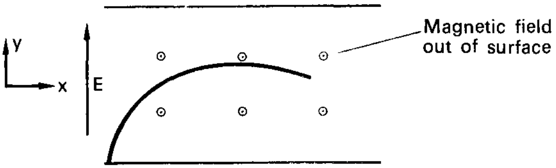


FIG. 12.16. Planar diode with magnetic field.

The equation of motion for the electron under these conditions, as shown in Fig. 12.16, are

$$Ee-Bev_x = mf_y = m\frac{dv_y}{dt} \tag{2.6.1}$$

and

$$Bev_y = mf_x = m\frac{dv_x}{dt}. \tag{2.6.2}$$

Differentiating eqn. (2.6.1) and substituting from eqn. (2.6.2)

$$\frac{d^2v_y}{dt^2} = -\left(\frac{Be}{m}\right)^2 v_y.$$

The solution of this equation is

$$v_y = K\sin\left(\frac{Be}{m}\right)t.$$

Also from eqn. (2.6.1),

$$Bev_x = E_e - m\,\frac{dv_y}{dt}$$

$$= Ee-mK\left(\frac{Be}{m}\right)\cos\left(\frac{Be}{m}\right)t. \tag{2.6.3}$$

At $t = 0$, $v_x = 0.$

Therefore $0 = Ee - KBe,$

i.e. $K = \frac{E}{B}.$

Therefore $v_y = \frac{E}{B} \sin\left(\frac{Be}{m}\right)t,$

i.e. $\frac{dy}{dt} = \frac{E}{B} \sin\left(\frac{Be}{m}\right)t.$

Integrating, $y = -\frac{E}{B}\left(\frac{m}{Be}\right) \cos\left(\frac{Be}{m}\right)t + K_1.$

At $t = 0$, $y = 0.$

Therefore $K_1 = \frac{Em}{B^2e},$

i.e. $y = \frac{Em}{B^2e}\left[1 - \cos\left(\frac{Be}{m}\right)t\right].$

Also from eqn. (2.6.3),

$$v_x = \frac{dx}{dt} = \frac{E}{B}\left[1 - \cos\left(\frac{Be}{m}\right)t\right],$$

i.e. $x = \frac{E}{B}\left[t - \left(\frac{m}{Be}\right) \sin\left(\frac{Be}{m}\right)t\right] + K_2.$

At $t = 0$, $x = 0.$

Therefore $K_2 = 0,$

i.e. $x = \frac{E}{B}\left[t - \left(\frac{m}{Be}\right) \sin\left(\frac{Be}{m}\right)t\right].$

From the expressions for x and y it follows that the path of the electron is a curve rising to a maximum value of $(2Em)/(B^2e)$ after a time $(\pi m)/(Be)$ and returns to $y = 0$ in a time $(2\pi m)/(Be)$ when the horizontal displacement is $(2\pi mE)/(B^2e)$.

The critical value of the magnetic field, i.e. when the electron just reaches the top electrode, is given by

$$d = \frac{2(V/d)m}{B^2 e},$$

where V is the direct voltage between the electrodes and d is the distance between them,

i.e. $$B = \frac{1}{d}\sqrt{\left(\frac{2mV}{e}\right)}.$$

This corresponds to a time t,

where $$t = \frac{\pi m}{e} d \sqrt{\left(\frac{e}{2mV}\right)}$$
$$= \pi d \sqrt{\left(\frac{m}{2eV}\right)}.$$

12.6.3. Coaxial diode with transverse magnetic field

Critical conditions for the coaxial diode shown in Fig. 12.17 may be derived by considering the energy equation and the expression relating torque and the change in angular momentum. The latter may be stated as

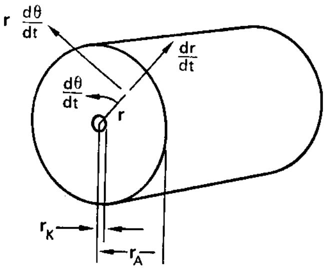


FIG. 12.17. Coaxial diode with magnetic field.

torque = rate of change of angular momentum,

i.e. $$\text{tangential force} \times \text{radius} = \frac{d}{dt}(I\omega),$$

where I is the moment of inertia and ω is the angular velocity,

i.e. $$Be\frac{dr}{dt}r = \frac{d}{dt}\left(mr^2\frac{d\theta}{dt}\right).$$

Integrating, and $$Be\frac{r^2}{2} = mr^2\frac{d\theta}{dt}+K.$$

If the electron leaves the cathode at right angles, at $r = r_K$,

$$\frac{d\theta}{dt} = 0 \quad \text{and} \quad K = \frac{Ber_K^2}{2}.$$

Usually r_K is small compared with r and K may be ignored,

i.e. $$\frac{Ber^2}{2} = mr^2\frac{d\theta}{dt}$$

or $$\frac{d\theta}{dt} = \frac{Be}{2m}. \tag{2.6.4}$$

Also equating the gain of kinetic energy to the loss of potential energy,

$$\frac{1}{2}m\left(\frac{dr}{dt}\right)^2+\frac{1}{2}mr^2\left(\frac{d\theta}{dt}\right)^2 = eV_r,$$

where V_r is the voltage at a radius r,

i.e. $$\frac{1}{2}m\left(\frac{dr}{dt}\right)^2+\frac{1}{2}mr^2\left(\frac{Be}{2m}\right)^2 = eV_r.$$

The condition for zero radial velocity (i.e. $dr/dt = 0$) is then

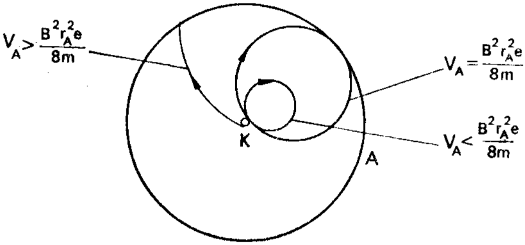
$$\frac{B^2r^2e}{8m} = V_r.$$

Under critical conditions, when the electron moves in a path such that it just reaches the anode, V_r is the applied voltage V_A and r the anode radius r_A.

Then

$$V_A = \frac{B^2 r_A^2 e}{8m}.$$

The electron paths for different values of V_A are shown in Fig. 12.18.



$V_A > \frac{B^2 r_A^2 e}{8m}$
$V_A = \frac{B^2 r_A^2 e}{8m}$
$V_A < \frac{B^2 r_A^2 e}{8m}$
K
A

Fig. 12.18. Electron paths in coaxial diode with magnetic field.

The time of transit for the critical condition may be obtained by integrating eqn. (2.6.4),

i.e.
$$\int_0^{\pi} \left(\frac{d\theta}{dt}\right) dt = \int_0^{t} \left(\frac{Be}{2m}\right) dt,$$

i.e.
$$t = \frac{2\pi m}{Be}.$$

In this ideal condition, for a given anode voltage, the anode current would be at its maximum value for magnetic field strengths less than the critical value and zero for field strengths greater than the critical value. In a practical diode some blurring would occur due to the effects of emission velocity, electron interaction, and asymmetrical fields. These cases are illustrated by Fig. 12.19.

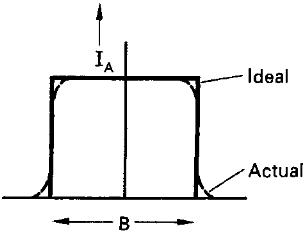


FIG. 12.19. Variation of anode current with magnetic field.

12.6.4. *The cavity magnetron*

The practical application of a diode with an axial magnetic field is the cavity magnetron illustrated in Fig. 12.20. The magnetic field *B* is adjusted to its critical value and, under favourable conditions, high-frequency oscillations will be set up. The superimposed alternating electric field distribution at a particular instant is shown in the figure.

Emitted electrons now come under the influence of the coaxial magnetic field, the direct electric field from anode to cathode, and the alternating electric field between the segments of the anode.

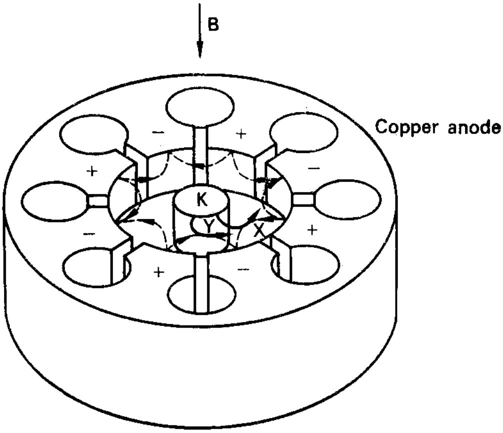


FIG. 12.20. Cavity magnetron.

For the instant illustrated by Fig. 12.20, electrons at X will be accelerated by the alternating field and are quickly returned to the cathode. Such electrons absorb energy from the alternating field.

Of more interest are electrons at the point Y. These are retarded by the alternating field giving up energy to that field. The reduction in velocity, and corresponding drop in the force due to the magnetic field, means that the electrons stay a relatively long time in this region constantly giving up energy. They eventually reach the anode as indicated by Fig. 12.21, but the energy they give up far exceeds that absorbed by electrons returning to the cathode, and the net effect is that oscillations once initiated will be maintained.

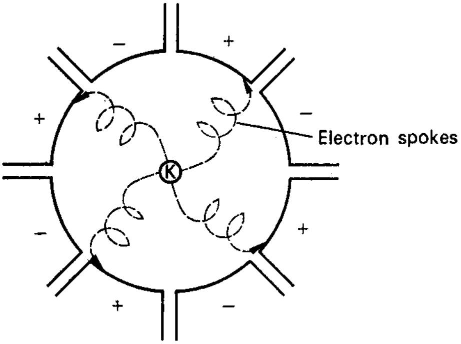


FIG. 12.21. Electron spokes.

In such a magnetron the eight cavities each form a tuned circuit and the alternating field effectively rotates giving a circular motion to the electron "spokes". Magnetic coupling to one of the cavities provides the output power.

Typically cavity magnetrons are operated under pulse conditions with a very low duty cycle (1 %). Pulses 1 μs long of peak value 100 kW may be produced with an overall efficiency of 70 %. The actual frequency of oscillation is usually in the range 500–10,000 MHz.

12.7. Focusing Effect of a Magnetic Field

As an alternative to electrostatic focusing of an electron beam, electromagnetic focusing may be employed. Ideally, a uniform axial magnetic field is desired, but in practice a thin magnetic lens, with its inherent non-uniform field, is incorporated.

12.7.1. Focusing effect of a uniform field

This topic has already been covered in Volume 1 but will be repeated briefly here. Consider a beam of electrons entering a uniform magnetic field as shown in Fig. 12.22 with a velocity $v = \sqrt{[(2eV_A)/m]}$, where V_A is the voltage of a previous accelerating electrode. Although generally moving in the z direction, some will be diverging at some angle θ, θ itself being small.

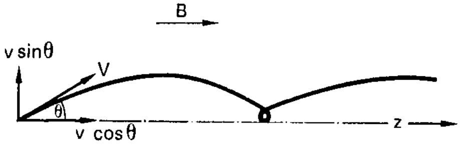


FIG. 12.22. Electromagnetic focusing.

Electrons then have an axial velocity $v \cos\theta$ and a radial velocity $v \sin\theta$. As indicated previously, the component $v \sin\theta$ gives rise to a circular motion in a plane perpendicular to the z axis. This has a radius given by

$$r = \frac{mv \sin\theta}{Be} = \frac{1}{B}\left(\sqrt{\frac{2mV_A}{e}}\right)\sin\theta.$$

At the same time the electron is moving along the z axis with a velocity $v \cos\theta$ which for small values of θ may be assumed to be v.

Hence the electron moves in a helix of radius $\frac{1}{B}\left(\sqrt{\frac{2mV_A}{e}}\right)\sin\theta$

with a pitch p of vt, where t is the time of one revolution,

i.e.
$$p = \left(\sqrt{\frac{2eV_A}{m}}\right)\frac{2\pi mv \sin\theta}{Bev\sin\theta}$$
$$= \frac{2\pi}{B}\sqrt{\left(\frac{2mV_A}{e}\right)}.$$

Since θ does not appear in this expression it follows that if θ is small a point focus will be produced at distances p, $2p$, etc., from the beginning of the magnetic field. If the magnetic flux density B is large, the electrons will be focused into a very tight beam. This is a requirement for devices such as the travelling wave tube.

12.7.2. *The thin magnetic lens*

Consider first the elemental magnetic lens shown in Fig. 12.23. This is analogous to the elemental electric lens discussed in section 12.4.2.

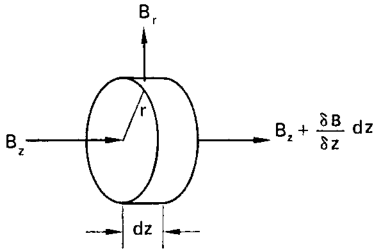


FIG. 12.23. Elemental magnetic lens.

The total magnetic flux entering the element must equal that leaving it.

Hence
$$\pi r^2 B_z - 2\pi r\, dz B_r - \pi r^2\left(B_z + \frac{\partial B_z}{\partial z}\, dz\right) = 0$$

or
$$B_r = -\frac{r}{2}\frac{\partial B_z}{\partial z}. \qquad (12.7.1)$$

In a practical magnetic focusing system as shown in Fig. 12.24, the lens may be regarded as extending from the plane at *A* to that at *B*. An electron at any point *P* is subject to three distinct forces due to this field, two acting tangentially but in opposite sense, and a third acting radially. The resultant tangential force F_ϕ on the electron is given by

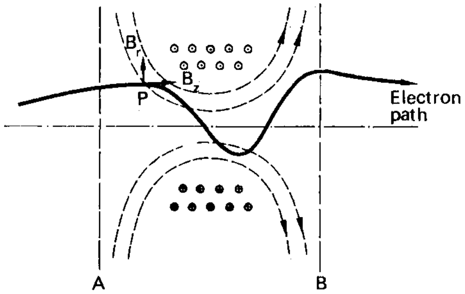


FIG. 12.24. Magnetic lens.

$$F_\phi = B_r e v_z - B_z e v_r = B_r e \frac{\partial z}{\partial t} - B_z e \frac{\partial r}{\partial t},$$

where B_r and B_z are the radial and axial components of the magnetic field, e is the charge on an electron, and v_z and v_r are the axial and radial components of velocity. But from eqn. (12.7.1),

$$B_r = -\frac{r}{2}\frac{\partial B_z}{\partial z}.$$

Therefore

$$F_\phi = -e\left(\frac{r}{2}\frac{\partial z}{\partial t}\frac{\partial B_z}{\partial z} + B_z\frac{\partial r}{\partial t}\right)$$

$$= -e\left(\frac{r}{2}\frac{\partial B_z}{\partial t} + B_z\frac{\partial r}{\partial t}\right).$$

Therefore
$$\begin{aligned}\text{torque} &= rF_\phi\\ &= -e\left(\frac{r^2}{2}\frac{\partial B_z}{\partial t}+rB_z\frac{\partial r}{\partial t}\right)\\ &= -\frac{e}{2}\frac{d}{dt}(r^2B_z).\end{aligned}$$

Hence equating torque to the rate of change of angular momentum,

$$-\frac{e}{2}\frac{d}{dt}(r^2B_z) = \frac{d}{dt}(mr^2\omega).$$

If the origin (i.e. $z = 0$) is taken at plane A at any distance, this expression may be integrated, giving

$$-\frac{e}{2}r^2B_{z_A}+\frac{e}{2}r^2B_z = mr^2\omega_A - mr^2\omega.$$

But at $z = 0$, $\quad B_{z_A} = 0 \quad$ and $\quad \omega_A = 0$.

Then at any distance z, $\quad \omega = -\dfrac{eB_z}{2m}$.

This shows that at any point the angular velocity is proportional to the axial component of the magnetic field and is zero when the electron leaves the magnetic lens. The negative sign follows from B_r, actually acting in the opposite direction to that indicated.

Allowing for the centrifugal force $mr\omega^2$, the radial force F_r on the rotating electron at any distance z is given by

$$F_r = B_z er\omega + mr\omega^2.$$

But $\omega = -\dfrac{eB_z}{2m}$.

Therefore
$$\begin{aligned}F_r &= -\frac{B_z^2e^2r}{2m}+mr\frac{e^2B_z^2}{4m^2}\\ &= -\frac{B_z^2e^2r}{4m}.\end{aligned}$$

This force is proportional to the radius and the negative sign shows it is always directed towards the axis. Hence a thin magnetic lens will produce a point focus.

Questions on Chapter 12

Q.1. Show that theoretically an electron reaches the speed of light when accelerated through a potential of 275 kV. What reduction takes place when relativistic effects are considered?

Q.2. Show that if an electron enters the region between two plane parallel equipotential surfaces at an angle θ, it will describe a parabolic path between these surfaces.

Q.3. Explain the action of a three-electrode lens when the centre electrode is at a higher potential than the outside electrodes.

Q.4. The buncher of a klystron is 1 kV above the cathode. If a sinusoidally alternating potential of 200 V (peak) is applied between the closely spaced grids of the buncher, estimate the shortest distance at which maximum bunching will occur. The frequency is 1 GHz and e/m is $1 \cdot 76 \times 10^{11}$ C/kg.

Q.5. A coaxial diode has a transverse magnetic field, its value being adjusted such that the anode current just flows. Determine the value of this field if the diameter of the anode is 1 cm and the anode voltage is 600 V.

Q.6. Show that the transit time of electrons under space-charge conditions is two-thirds of that under temperature-limited conditions.

Appendix. Table of Laplace Transforms

$$F(s) = \int_0^\infty e^{-st} f(t)\,dt$$

$f(t)$	$F(s)$
1	$\frac{1}{s}$
$e^{\alpha t}$	$\frac{1}{s-\alpha}$
$\sin \alpha t$	$\frac{\alpha}{s^2+\alpha^2}$
$\cos \alpha t$	$\frac{s}{s^2+\alpha^2}$
t	$\frac{1}{s^2}$
$\sinh \alpha t$	$\frac{\alpha}{s^2-\alpha^2}$
$\cosh \alpha t$	$\frac{s}{s^2-\alpha^2}$
$e^{-\alpha t}\sin \beta t$	$\frac{\alpha}{(s+\alpha)^2+\beta^2}$
$e^{-\alpha t}\cos \beta t$	$\frac{s+\alpha}{(s+\alpha)^2+\beta^2}$
$te^{-\alpha t}$	$\frac{1}{(s+\alpha)^2}$
$e^{-\alpha t}\sinh \beta t$	$\frac{\beta}{(s+\alpha)^2-\beta^2}$
$e^{-\alpha t}\cosh \beta t$	$\frac{s+\alpha}{(s+\alpha)^2-\beta^2}$
$\frac{dx}{dt}$	$s\bar{x}-x_0$
$\frac{d^2x}{dt^2}$	$s^2\bar{x}-sx_0-x_0'$

Answers to Numerical Examples

CHAPTER 1

A1. $VC(1-\varepsilon^{-t/RC})$.
A4. $C = 500$ pF; $L = 12{\cdot}5$ mH.
A5. Series capacitances: 332 pF; 278 pF; 348 pF.
Parallel circuits: 20 μH; 498 pF; 6 μH; 6·55 μH; 6080 pF; 20 μH; 498 pF.
A6. Series arms: 3·18 μH; 0·2 μF.
Shunt arm: 3·58 mH; 0·177 μF.

CHAPTER 2

A1. 0·265 Å; −54·8 eV; 1·058 Å; −13·7 eV.
A2. 0·00568 m^2/Vs; $5{\cdot}68\times10^{-11}$ m^3/C; 1·76 A/cm^2.
A3. 0·279 Ωm.
A5. 0·0524 m^3/C; 0·39 m^2/Vs.

CHAPTER 5

A1. (a) 1·5 kΩ; (b) 10·3 kΩ; (c) −42·4 times; (d) 63·6 times.
A3. $y_{ie} = y_{ib}+y_{rb}+y_{ob}+y_{fb}$,
$y_{re} = -(y_{rb}+y_{ob})$,
$y_{fe} = -(y_{ob}+y_{fb})$
$y_{oe} = y_{ob}$.
A4. $n = 1$, $R_N = 312\ \Omega$, $C_N = 6{\cdot}5$ pF.
A5. Common base: $Z_{\text{in}} = r_e+r_b(1-\alpha)$; $Z_{\text{out}} = r_c-\dfrac{\alpha r_c r_b}{r_e+r_b+Z_s}$.
Common collector: $Z_{\text{in}} = r_b+\dfrac{R_L}{(1-\alpha)}$; $Z_{\text{out}} = r_e+\dfrac{(R_s+r_b)}{(1-\alpha)}$.
A6. $Z_{\text{in}} = 2{\cdot}3$ kΩ; $Z_{\text{out}} = 55$ kΩ.

CHAPTER 6

A1. 1700 pC.
A3. $T_o \log_\varepsilon\left[\dfrac{I_B+I_{B_1}}{I_c/(h_{FE}+I_B)}\right]$.

CHAPTER 8

A2. 16·2 μV.
A3. 6 stages.
A5. 2·4 times.

CHAPTER 9

A2. $Z_m\left[1-\frac{\mu(R+\beta Z)}{Z_{\text{out}}+Z}\right]$.
A3. $-\frac{\mu(\mu R+r_a)}{R(1+\mu)+2r_a}$.
A4. $0{\cdot}48+j3{\cdot}14$ mS.

CHAPTER 10

A2. 2 s.
A3. $AR = A_L R_C$.

CHAPTER 11

A4. $(\bar{A}\cdot C+\bar{B})$.
A5. $\bar{B}\cdot C+B\cdot\bar{A}$.

CHAPTER 12

A1. 26%.
A4. 4·7 cm.
A5. 28·7 mT.

Index

Amplifier
  common-gate 238
  common source 238
  direct-coupled FET 239
  distributed 262
  r.f. FET 240
AND circuit 280
Applegate diagram 328
Atom 60
Attenuation bond 10
Attenuation coefficient 10

Base time constant 172
Bohr atom 60
Boolean algebra 281
Bootstrap sweep 276
Buried-*n* layer 195

Cascade circuit
  FET 258
  transistor 257
  valve 256
Cathode lead inductance 259
Characteristic admittance, π network 9
Characteristic impedance, T network 8
  band pass filter 27
  band stop filter 33
  high pass filter 19
  low pass filter 13
Charge control 167
Coaxial line oscillator 236
Collector time constant 171
Correlation factor 221
Cut-off frequency
  band pass filter 24
  band stop filter 31
  high pass filter 17
  low pass filter 12

Deflexion
  electric 318
  magnetic 332
Delay time 176
Density of states 71
Depletion region 89
Design impedance
  band pass filter 27
  band stop filter 35
  high pass filter 21
  low pass filter 15
Diffusion 188
Diffusion current 77
Diffusion length 91
Displacement error 267
Drift velocity electron 66
  hole 69
Dynamic deflexional sensitivity 321

Ebers-Moll relations 96
Einstein relation 79
Electric lens 324
Electron orbits 60
Electron shells 62
Energy barrier (*pn* junction) 89
Epitaxial deposition 189
Equivalent circuit
  FET 127
  "h" 130
  hybrid-π 155
  "T" 147
  "y" 138

Equivalent noise resistance 218
Evaporation 190
Excess base charge 173
Exclusive OR circuit 290

Fall time 181
Fan in 299
Fan out 299
Feedback 245
Fermi–Dirac distribution law 73
Fermi level
  intrinsic semiconductor 76
  n-type semiconductor 83
  p-type semiconductor 85
Field effect transistor 103
Filter 8
  m-derived 36
  m-derived band pass 42
  m-derived band stop 46
  m-derived high pass 40
  m-derived low pass 38
  m-derived half section 50
    band pass 55
    band stop 56
    high pass 54
    low pass 53
  prototype 11
    band pass 22
    band stop 28
    high pass 16
    low pass 11
Full adder 292

Gauss's theorem 312

Half adder 291
Hall angle 68
Hall coefficient 68
Hall effect
  conductors 67
  in extrinsic semiconductors 87

Inclusive OR circuit 283
Insulated gate, FET
  depletion type 112
  enhancement type 116
  enhancement/depletion type 126

Junction, FET 105

Klystron 327

Laplace transform 1
Law of mass action 81
Logic
  complementary transistor 306
  current mode 307
  diode–transistor 304
  direct-coupled transistor 302
  emitter-coupled 307
  modified diode–transistor 304
  resistor–capacitor–transistor 303
  resistor–transistor 303
  transistor–transistor 305

Magnetic lens 343
Magnetron 340
Mask 186
Measurement of noise factor 230
Miller integrator 269
Mobility 66
Monolithic circuit 184
MOST 103
Multivibrator 265

NAND circuit 286
Noise 210
  bandwidth 211
  factor 225
  figure 225
  flicker 215
  fundamental 210

Noise *(cont.)*
  Johnson 210
  margin 299
  resistance 210
  shot 214
  temperature 232
  thermal 210
  white 210
NOR circuit 287
NOT circuit 285

Oxidation 185

Pass band 10
Phase change coefficient 10
Photo-engraving 185
Photo-etching 187
Photo resist 187
Pinch-off voltage 112
Planar magnetron 335
pn junction 89
Poisson's equation 109, 313
Propagation coefficient 8
Propagation delay 301

Quantum numbers 62
  magnetic 63
  principal 62
  spin 63
  subsidiary 63

Ramp function 266
Reflex klystron 330
Refraction (by electric field) 323
Relativistic effects 310
Rise time 178
Rubylith 186

Saturation time constant 173
Signal-to-noise ratio 222
Slope error 266
Source follower 238
Static deflexional sensitivity 320
Storage time 181
Switch-off charge 174
Switch-on charge 173

Three-halves power law 314, 316
Transistor
  diffused collector 192
  epitaxial 194
  triple-diffused 194
Transit time 260
Transmission error 267

Velocity modulation 327

Wien bridge oscillator 254